

Zwei Bereiche aus dem Genom
des *Cercopithecus aethiops*:
Das c-myc Gen und ein Abschnitt
mit Sequenzhomologie
zu Transkriptionskontrollregionen

Herbert Kühnel



Digitized by the Internet Archive
in 2026

**Zwei Bereiche aus dem Genom des
Cercopithecus aethiops:**

**Das c-myc Gen
und ein Abschnitt mit
Sequenzhomologie zu
Transkriptionskontrollregionen**

Dissertation der Fakultät für Chemie und Pharmazie der
Ludwig-Maximilians-Universität München

vorgelegt von
Herbert Kühnel

München

1986

1. Gutachter: Prof. Dr. E.-L. Winnacker
2. Gutachter: Prof. Dr. G. Hartmann

Tag der mündlichen Prüfung: 28.11.1986

Vorliegende Arbeit wurde auf Anregung und unter Anleitung von Herrn Prof. Dr. E.-L. Winnacker am Institut für Biochemie der Ludwig-Maximilians-Universität München ausgeführt.

Herrn Prof. Dr. E.-L. Winnacker danke ich für die Möglichkeit zum freien, selbständigen Arbeiten in seinem bestens ausgestatteten Labor.

Herrn Dr. Martin Lipp danke ich für sein Engagement für den Arbeitskreis, für sein Interesse an meiner Arbeit und dafür, daß er myc nach München gebracht hat.

Frau Dr. Inge Kruczek danke ich für ihr Interesse an meiner Arbeit und dafür, daß sie p9 nach München gebracht hat.

Herrn Peter Hartl danke ich für die Zusammenarbeit bei der Präparation von Klonierungsvektoren.

Herrn Peter Heinrich danke ich für seine Einführung in die Technik des Sequenzierens.

Ich danke Herrn Dr. Gerhard Greull, ohne den radioaktives Arbeiten an diesem Insitut nicht möglich wäre.

Ich danke allen Mitarbeitern, die zu einer guten Atmosphäre im Insitut beigetragen haben.

INHALTSVERZEICHNIS

1.	EINLEITUNG	1
1.1.	Der Bakteriophage Lambda 1059 als Klonierungs- vektor für Genbanken	1
1.2.	myc	3
1.3.	p9, eine Integrationsstelle der Adl2 DNA ins Genom der Hamster Tumorzelllinie CLAC1	6
1.4.	Aufgabenstellung	10
2.	MATERIAL	11
2.1.	Zellen, Phagen, Plasmide	11
2.2.	Enzyme, Chemikalien, Materialien, Geräte	12
2.3.	Medien, Lösungen, Puffer	16
3.	METHODEN	18
3.1.	Zellen, Bakterien, Phagen	18
3.1.1.	Züchtung von TC7 Zellen	18
3.1.2.	Züchtung von Bakterien	18
3.1.3.	Züchtung von Phagen	19
3.1.3.1.	Vermehrung von Lambda Phagen auf Agarplatten	19
3.1.3.2.	Titration	19
3.1.3.3.	Lambda Phagenstock anlegen	20
3.1.3.4.	Plaque Reinigung	20
3.2.	Grundtechniken	20
3.2.1.	Phenolextraktion von Nukleinsäure Lösungen	20
3.2.2.	Fällen von Nukleinsäuren	21
3.2.3.	Konzentrationsbestimmung von Nukleinsäure Lösungen	21
3.2.4.	Spaltung von DNA durch Restriktionsendonukleasen	21
3.2.5.	Gelelektrophorese von Nukleinsäuren	22
3.2.5.1.	DNA	22
3.2.5.2.	RNA	23

3.2.6.	Isolierung von DNA Fragmenten aus Agarose Gelen	24
3.3.	Isolierung von Nukleinsäuren	24
3.3.1.	Isolierung von hochmolekularer DNA aus Gewebe- kultur Zellen	24
3.3.2.	Isolierung von cytoplasmatischer RNA aus Gewebe- kultur Zellen	25
3.3.3.	Isolierung polyadenylierter RNA	25
3.3.4.	Isolierung von Plasmid DNA	26
3.3.4.1.	CsCl Gleichgewichtszentrifugationsmethode für größere Mengen	26
3.3.4.2.	Schnellaufschluß Methode für kleine Mengen	27
3.3.5.	Isolierung von Lambda DNA	28
3.3.5.1.	CsCl Gleichgewichtszentrifugationsmethode für größere Mengen	28
3.3.5.2.	Schnellpräparation für kleine Mengen	29
3.3.6.	Isolierung von M13 DNA	29
3.3.6.1.	Schnellaufschluß Methode zur Isolierung von M13 ssDNA	29
3.3.6.2.	Isolierung von M13 RF DNA	30
3.4.	Vorbereitung der DNA für die Klonierung	31
3.4.1.	Präparation von Fragmenten zellulärer DNA eines bestimmten Größenbereichs	31
3.4.2.	Präparation von Lambda Vektorarmen mit BamHI Enden	31
3.4.3.	Präparation von Plasmid- oder M13 Vektoren	32
3.5.	Klonierungsschritte	33
3.5.1.	Ligase Reaktionen	33
3.5.1.1.	Verknüpfung von Lambda Phagenarmen mit Fragmenten zellulärer DNA	33
3.5.1.2.	Verknüpfung von Plasmid- oder M13 Vektoren mit der zu klonierenden DNA	33
3.5.2.	Transformation von Bakterienzellen	34
3.5.2.1.	Präparation transformationskompetenter Bakterien (<i>E.coli</i> HB101)	34

3.5.2.2.	Transformation von <i>E.coli</i> HB101	34
3.5.2.3.	Transformation von <i>E.coli</i> JM103 durch M13 RF DNA	35
3.5.3.	<i>In vitro</i> Verpackung von Lambda DNA	35
3.6.	Charakterisierung von Nukleinsäure Molekülen durch Hybridisierung	37
3.6.1.	Markierung von DNA durch Einbau radioaktiver Nukleotide an Einzelstrang Brüchen (Nick Translation)	37
3.6.2.	Übertragung von DNA oder RNA auf Membranfilter (Southern- bzw. Northern Blotting)	38
3.6.3.	Regenerierung bereits hybridisierter Membranfilter	39
3.6.4.	Hybridisierung von filtergebundener DNA oder RNA (Southern- bzw. Northern Blot Hybridisierung)	39
3.6.5.	Isolierung spezifischer Klone aus einer Genbank in Lambda Phagen	40
3.6.6.	Komplementaritätstest von M13 Inserts	41
3.7.	DNA Sequenzierung nach der Kettenabbruch Methode	42
3.7.1.	Vorbereitung der Probe (Template/Primer)	42
3.7.2.	Sequenzierungsreaktion	43
3.7.3.	Gel Elektrophorese	44
4.	ERGEBNISSE	45
4.1.	Genbank der TC7 Zelllinie	45
4.2.	Das c-myc Gen der TC7 Zelllinie	47
4.2.1.	Isolierung und Restriktionskartierung	47
4.2.2.	Subklonierungen	50
4.2.3.	Sequenzierung	51
4.2.4.	Southern- und Northern Blot Analyse	54
4.2.5.	Sequenzvergleich	56
4.2.6.	Potentielle regulatorische Elemente des c-myc Gens	65
4.2.6.1.	Das Sequenzmotiv GTTACAT	65
4.2.6.2.	Die repetitive Polypyrimidin Sequenz	67

4.2.6.3	Die Konsensussequenz der Spl Bindungsstelle	69
4.2.6.4.	Der zum hsp83 HSE homologe Bereich	71
4.2.6.5.	Die vollständig konservierte Sequenz nach der cap-Stelle 2	74
4.3.	p9 homologe DNA des TC7 Genoms	75
4.3.1.	Isolierung	75
4.3.2.	Charakterisierung des L452 Subklons pHK6.1	79
4.3.3.	Sequenzierung des 1.13kb PstI Fragments	82
4.3.4	Das p9 Hybridisierungsphänomen	84
4.3.5.	Charakterisierung des L457 Subklons pHK5.9	86
4.3.6.	Genomische Analyse der TC7 Zelle bezüglich p9 und pHK6.1	87
4.3.7.	Sequenzvergleichsanalyse	89
4.3.7.1.	Sequenzhomologien zwischen TC7p9 und p9	89
4.3.7.2.	GC ₃ -Cluster	90
4.3.7.3.	Gemeinsamkeiten von p9, TC7p9, den pCaOri Klonen und der regulatorischen Region der SV40 DNA	92
4.3.7.4.	Charakterisierung der TC7p9 Sequenz	99
4.3.7.5.	Charakterisierung der p9 (HTC) Sequenz	102
4.3.7.6.	Sequenzen mit alternierenden Purin-/Pyrimidin- nukleotiden	106
4.3.7.7.	Offene Leserahmen von TC7p9	110
4.3.7.8.	Datenbank Sequenzen mit Homologie zu TC7p9 oder p9	116
4.3.7.9.	Homologie von TC7p9 mit Sequenzen der Immun- globulin Gene	117
4.3.7.10.	Homologie zwischen p9 und einem X.laevis rRNA Gen	119
4.3.8.	Hypothese zur Erklärung des Hybridisierungs- phänomens	121
5.	DISKUSSION	127
5.1	Potentielle regulatorische Elemente des c-myc Gens	127

5.2.	p9 und TC7p9, zwei weitere zelluläre Sequenzen mit Homologie zum regulatorischen Bereich der SV40 DNA	130
6.	ZUSAMMENFASSUNG	134
7.	ANHANG	135
7.1.	Zum Vergleich herangezogene Sequenzen	135
7.1.1.	c-myc Gen des Huhns	135
7.1.2.	N-myc, Exon 1	136
7.1.3.	Regulatorische Region der SV40 DNA	137
7.1.4.	Hamster DNA Anteil von p9	138
7.1.5.	pCaOri5.01	138
7.1.6.	pCaOri7.01	139
7.1.7.	pCaOri9.32	139
7.2.	DNA Sequenzelemente	140
7.2.1.	Adenovirus Enhancer Stammsequenzen	140
7.2.2.	SV40 Enhancer Stammsequenzen	141
7.2.3.	SV40 SphI Motive	142
7.2.4.	TCCCCA Motive	142
7.2.5.	Histon Gen Enhancerelemente	143
7.2.6.	Regulatorische Elemente der hsp Gene	143
7.2.7.	Weitere Sequenzelemente	144
7.2.8.	Spleiß Signale	145
7.2.9.	S1 Nuklease hypersensitive Stellen	146
7.3.	Alternierende P/Q Sequenzen	147
7.4.	Statistische Signifikanz von Sequenzhomologien	151
8.	LITERATURHINWEISE	154

ABKÜRZUNGEN

Ad	Adenovirus
bp	Basenpaar(e)
CAT	Chloramphenicol Acetyltransferase
cpm	gezählte Zerfälle pro Minute
dsDNA	doppelsträngige DNA
EBV	Epstein-Barr Virus
HSE	Hitzeschock Element
hsp	Hitzeschock Protein (Gen)
HSV	Herpes Simplex Virus
IE	immediate early
kb	Kilobasenpaar(e)
L	Liter, zur Unterscheidung von l (eins)
N	A, C, G oder T
P	A oder G (*)
pfu	Plaques bildende Einheiten
Pos.	Position(en)
Q	C oder T (*)
R	A oder T (*)
RF DNA	replizierende Form der (M13) DNA
rpm	Umdrehungen pro Minute
RT	Raumtemperatur
ssDNA	einzelsträngige DNA
V	A oder C (*)
W	C oder T (*)

(Die Abkürzungen einiger Chemikalien sind in Absatz 2.2 angegeben.)

(*) wegen der Verwendung der Microgenie™ Programme (Queen & Korn, 1984). Die Symbole stimmen nicht mit der von der NC-IUB (1985) empfohlenen Nomenklatur überein.

1.

EINLEITUNG

1.1.

Der Bakteriophage Lambda 1059
als Klonierungsvektor für
Genbanken

Das von Karn et al. (1980) entwickelte Bakteriophagenderivat Lambda 1059 ist ein BamHI Substitutionsvektor, der sich für das Anlegen von Genbanken eignet. Die Lambda 1059 DNA kann durch BamHI in 3 Teile gespalten werden: in das zentrale, nicht essentielle 14kb Fragment und die beiden "Arme", das linke 19.6kb- und das rechte 9.2kb Endfragment. Die über die kohäsiven, 5'-überstehenden 12 Nukleotide der Lambda cos-Stelle verbundenen Arme lassen sich abtrennen. Ligiert man Lambda Arme mit 9-23 kb langen Fragmenten fremder DNA, so entstehen, nach *in vitro* Verpackung, lebensfähige rekombinante Phagen.

Bietet man als Insert die (durch unvollständige Spaltung mit Sau3A oder MboI) in entsprechend lange Stücke gespaltene, genomische DNA einer Zelle an und erhält danach so viele rekombinante Phagen (bei Vertebraten mehr als 5×10^5), daß diese die eingesetzte DNA praktisch komplett enthalten, so spricht man von einer Genbank.

Die Anzahl (N) der Klone einer Genbank, die man untersuchen muß, um daraus mit einer bestimmten Wahrscheinlichkeit (P) jedes beliebige Gen zu isolieren, hängt vom Größenverhältnis (f) der Insertionen zum Genom ab:

$$N = \ln(1-P) / \ln(1-f); \text{ (Clarke \& Carbon, 1976).}$$

Das Lambda 1059 System bietet eine doppelte Selektion auf rekombinante DNA. Zum einen werden nur Lambda Arme verpackt, die ein Insert entsprechender Größe aufgenommen haben; zum anderen können auf P2-Lysogenen nur rekombinante Phagen wach-

sen. P2-Lysogene sind *E.coli* Stämme (z.B. Q359), die das Genom des Phagen P2 im Chromosom integriert haben; P2 ist ein temperater *E.coli* Phage, der zu einer anderen Gruppe als Lambda gehört.

Das zentrale BamHI Fragment von Lambda 1059 trägt die Funktionen *red* (*exo* und *bet*) und *gam*, die den Phänotyp Spi⁺ (sensitive to P2 interference) vermitteln: die Unfähigkeit auf P2-Lysogenen zu wachsen. Ein rekombinanter Phage besitzt dagegen diese Funktionen nicht, hat den Phänotyp Spi⁻ und kann auf einem P2-Lysogen wachsen. Er wird aber dafür vom *recA* System des Wirts abhängig, denn ihm fehlt das *gam* Genprodukt, das verhindert, daß konkatemere Formen der Lambda DNA durch das *recBC* Genprodukt, einer Exonuklease des Wirts, abgebaut werden. So kann Lambda DNA nur noch ausgehend von multimeren, zirkulären Formen verpackt werden. Diese müssen, beim Fehlen der *red* Genprodukte des Phagen, durch das *recA* System des Wirts gebildet werden. Damit die *recA* abhängige Rekombination effizient abläuft, ist auf dem rechten Phagenarm die Mutation *chi3* (cross-over hot spot instigator) eingeführt worden.

Lambda 1059 hat eine weitere Besonderheit. Das zentrale BamHI Fragment trägt ein ColEI Replikon und zwei Antibiotika Resistenzen. Die Phagen DNA kann deshalb in Gegenwart eines Repressors wie ein Plasmid vermehrt werden ("Phasmid", Phage/Plasmid). Im Zusammenhang mit dem Durchmustern (screening) einer Genbank ist dabei zu beachten, daß die Lambda 1059 DNA mit jeder Probe hybridisiert, die ebenfalls entsprechende Plasmidsequenzen enthält. So kann man, etwa durch Hybridisierung mit pBR322, kontrollieren, ob und wenn, wieviel Ausgangsphage in der Genbank vorhanden ist, darf sich aber andererseits nicht täuschen lassen, indem man diese Signale für spezifisch hält.

Weiterentwickelte Phagenvektoren haben zusätzliche Schnittstellen, durch die man das Insert exakt wieder ausschneiden kann, aber keinen Plasmidanteil (Frischauf et al., 1983; Karn

et al., 1984).

Soweit es die Verwendung als Klonierungsvektor angeht, ist der Bakteriophage Lambda in Winnacker (1984) beschrieben; einen Gesamtüberblick bieten Hendrix et al. (1983).

1.2.

myc

Das c-myc Protoonkogen ist das zelluläre Homologe (Sheines & Bishop, 1979) zum v-myc Onkogen des Avian Myelocytomatosis Virus MC29 (Sequenz: Alitalo et al., 1983). MC29, ein defektes RNA-Tumovirus, besitzt das transformierende Δ gag-myc Gen mit dem Genprodukt P110^{gag-myc} (Übersichtsartikel: Bishop, 1983). Die c-myc Gene von Mensch (Sequenz: Colby et al., 1983; Battey et al., 1983., Gazin et al., 1984), Maus (Sequenz: Stanton et al., 1983; Bernard et al., 1983) und Huhn (Sequenz: Watson et al., 1983a; Shih et al., 1983) sind bekannt. In *Drosophila* wurden myc verwandte Sequenzen durch Hybridisierung nachgewiesen (Shilo & Weinberg, 1981).

Das c-myc Protoonkogen hat 3 Exons. Exon 1 ist eine ungewöhnlich lange 5'-untranslatierte Region. Exon 2 und Exon 3 sind kodierend. Das c-myc Genprodukt ist ein Protein mit - bei Mensch und Maus - 439 Aminosäuren, das wie P110^{gag-myc} (Donner et al., 1982) im Zellkern (Hann et al., 1983) lokalisiert ist und mit hoher Affinität, sequenzunspezifisch an DNA bindet (Persson & Leder, 1984; Watt et al., 1985).

Ein durch Rearrangement aktiviertes c-myc Gen konnte in bestimmten Tumoren nachgewiesen werden und ist wahrscheinlich an deren Entstehung beteiligt:

- B-Zell Lymphome der Hühnervögel und T-Zell Lymphome der Nager, deren c-myc Gen durch Provirus Insertion verändert ist (Übersichtsartikel: Varmus, 1984).

- Burkitt Lymphome und akute B-Zell Leukämien des Menschen, die Chromosomentranslokationen aufweisen, durch die das c-myc Gen (liegt auf Chromosom 8) mit einem der 3 Immunglobulin Genorte (liegen auf den Chromosomen 2, 14 und 22) verbunden wird (Übersichtsartikel: Leder et al., 1983). Dabei kann der Bruchpunkt des Chromosoms 8 weit "vor" oder "hinter" c-myc, aber auch in der unmittelbaren 5'-untranskribierten Region, in Exon 1 oder Intron 1 liegen (Übersichtsartikel: Perry, 1983).
- Plasmacytome der Maus, die entsprechende Chromosomentranslokationen aufweisen (z.B. Crews et al., 1982).
- Verschiedene menschliche Tumoren, in deren Genom das c-myc Gen amplifiziert ist (z.B. Collins & Groudine, 1982).
- Die Zelllinie SKW-3 von einer menschlichen T-Zell Leukämie, die eine t(8;14)(q24;q11) Translokation aufweist. Dort liegt der Bruchpunkt des Chromosoms 8 nahe dem 3'-Ende des c-myc Gens und der Bruchpunkt des Chromosoms 14 zwischen der variablen und der konstanten Region des Gens für die α -Kette des T-Zell Rezeptors (Erikson et al., 1986; Shima et al., 1986).

Trotz intensivster Forschung ist die Funktion von c-myc in der normalen Zelle, bei der Tumorentstehung und in der transformierten Zelle erst ansatzweise aufgeklärt. Gleiches gilt für die Regulation der c-myc Genexpression. Zudem ist, oder scheint, vieles noch widersprüchlich. (Übersichtsartikel: Cole, 1985.)

c-myc wird in allen bislang untersuchten Zelltypen exprimiert (Westin et al., 1982; Müller et al., 1983; Stewart et al., 1984b). In der proliferierenden Zelle ist eine weit größere Menge an c-myc mRNA vorhanden als in einer differenzierten, ruhenden Zelle (Kelly et al., 1983). Dies wird anscheinend mehr durch posttranskriptionelle Regulation als durch Transkriptionskontrolle erreicht (Dani et al., 1985; Blanchard et al., 1985). Die c-myc mRNA ist mit einer Halbwertszeit von 10-15 min sehr instabil (Dani et al., 1984). Dies liegt zum Teil an Exon

1, denn fehlt dieses, ist die mRNA um den Faktor 4-5 stabiler (Piechaczik et al., 1985).

In den Fällen, in denen man die Herkunft der c-myc mRNA bestimmen konnte, zeigte sich, daß nur das durch Translokation aktivierte Allel (Taub et al., 1984b) oder nur das Transgen (Adams et al., 1985) transkribiert wird.

In Tumorzellen wird c-myc konstitutiv, jedoch nicht unbedingt stärker exprimiert als in normalen, sich teilenden Zellen (Keath et al., 1984). Das Fehlen von Exon 1 reicht für die Wirkung von c-myc in der Tumorentstehung nicht aus, auch nicht das Versetzen des kompletten Gens in eine andere chromosomale Umgebung (Adams et al., 1985). Ausreichend für die Aktivierung von c-myc zum transformierenden Onkogen ist dagegen die Wirkung eines benachbarten Enhancers (Land et al., 1983; Adams et al., 1985). Alles in allem geht man davon aus, daß c-myc durch Dereglulation zum Onkogen aktiviert wird (Übersichtsartikel: Klein & Klein, 1985).

Zusätzlich zur Aktivierung von c-myc ist mindestens ein weiteres Ereignis für eine Tumorentstehung nötig. Dies zeigte sich bei der Transformation von Zellen durch c-myc oder v-myc zusammen mit den Onkogenen ras, AdElb oder Py-mT (Land et al., 1983) und auch in der klonalen Natur der Tumoren c-myc transgener Mäuse (Stewart et al., 1984a; Adams et al., 1985). Man nimmt an, daß myc und andere Onkogene wie AdEla, Py-T, SV40-T, N-myc, myb und p53 die Immortalisierung einer Zelle bewirken können (Übersichtsartikel: Bishop, 1985; Weinberg, 1985).

Die Genprodukte von myc, myb und AdEla zeigen strukturelle Ähnlichkeiten. Die größte Homologie besteht dabei zwischen v-myc und Ela von Ad12 (Ralston & Bishop, 1983).

Die zellulären Onkogene N-myc und L-myc, die in bestimmten Tumoren amplifiziert sind, zeigen so große Homologie zu c-myc bzw. v-myc, daß man, wie die Namensgebung schon zeigt, von der

myc Gen Familie spricht. N-myc (Schwab et al., 1983; Kohl et al., 1983) hat die gleiche Genstruktur wie c-myc (Kohl et al., 1986) und kann, zusammen mit ras, Gewebekulturzellen transformieren (Yancopoulos et al., 1985). Über L-myc (Nau et al., 1985) gibt es erst wenige Daten. Die Arbeiten von Shilo & Weinberg (1981) und Dalla-Favera et al. (1982) lassen erwarten, daß weitere verwandte Gene gefunden werden.

1.3.

p9, eine Integrationsstelle der Adl2 DNA ins Genom der Hamster Tumorzelllinie CLAC1

Eine der Übergangsstellen vom linken Ende der integrierten Adl2 DNA ins Genom der Adl2 transformierten Hamster Zelllinie CLAC1, wurde als 6.4kb EcoRI Fragment charakterisiert (Stabel et al., 1980) und in Form eines rekombinanten Lambdaphagens isoliert. p9, ein Subklon davon, ein 757bp Sau3A Fragment in der BamHI Stelle von pBR322, wurde sequenziert (Stabel & Dörfler, 1982). p9 besteht aus 529 bp Hamster DNA und der Adl2 Sequenz (van Ormond & Galibert, 1984) von Pos. 175-402. Dies bedeutet, daß der Adl2 DNA an dieser Integrationsstelle die inverse terminale Repetition (ITR, Pos. 1-164) und weitere 10 bp fehlen.

Der nicht kodierende Bereich am linken Ende der Adenovirus DNA enthält essentielle virale Funktionen: Die ITR, mit dem Ursprungspunkt der Replikation (Übersichtsartikel: Fütterer & Winnacker, 1984), ein cis-Element für die Verpackung (Hammarskjöld & Winberg, 1980) und den Promotor der Ela Transkriptionseinheit.

Das Ela Gen der Adenoviren ist die Transkriptionseinheit, die im Verlauf der Infektion am frühesten (immediate early) expri-

miert wird (Übersichtsartikel: Levine, 1984). Der Ela Promotor muß folglich allein von zellulären Komponenten erkannt werden und mit zellulären Promotoren konkurrieren können. Er sollte einem zellulären Promotor wenigstens in einigen Bereichen gleichen. Das Protein der Ela 13S mRNA (Montell et al., 1982) induziert die anderen frühen Virusgene, kann aber auch bestimmte zelluläre Gene aktivieren (Übersichtsartikel: Kingston et al., 1985). Ela ist eines der beiden Onkogene der Adenoviren (Übersichtsartikel: van der Eb & Bernards, 1984).

Die 5'-Region des Ela Gens enthält Bereiche, die als Enhancer (Übersichtsartikel: Serfling et al., 1985) charakterisiert wurden:

Entfernt man aus der Ad5 DNA die Sequenz von 194-358 (entspricht dem Abschnitt von -305 bis -141 bezogen auf die Ela cap-Stelle bei Pos. 499), so sinkt die Transkription *in vivo* auf 5-7% des ursprünglichen Wertes, wird aber größtenteils wiederhergestellt, wenn man diesen Bereich 3' zur Ela kodierenden Region, egal in welcher Orientierung, insertiert. Diese Sequenz erhöht, wenn sie, wiederum unabhängig bezüglich Lage und Orientierung, auf einem Plasmid mit dem HSV Thymidinkinase Gen sitzt, die Zahl der stabilen Transformanten (Hearing & Shenk, 1983).

Wichtige Bestandteile dieser Enhancerregion (Ad Enhancer Stammsequenzen) sind die Nukleotidfolgen AGGAAGTGAAA bei Pos. 297 und AGGAAGTGACA bei Pos. 198. Dazu homologe Sequenzen befinden sich in nahezu gleicher Entfernung zur Ela cap-Stelle, nämlich bei "-200" bzw. "-300", auch in Adenoviren der Subgruppen A und B. Eine Ad Enhancer Stammsequenz enthalten auch andere virale Enhancer (siehe Anhang 7.2.).

Der Ad2 Bereich 1-270 enthält ein für die effiziente Transkription des Ela Gens wichtiges Element (Sassone-Corsi et al., 1983). Durch Transfektion entsprechender Plasmidkonstrukte und S1 Analyse konnte gezeigt werden (transient expression), daß

die Sequenz 149-178 (-349 bis -320) den Ad2 MLP (major late promoter) aktivieren kann, wenn man sie in dessen 5'-Bereich inseriert (Hen et al., 1983). Das auf diese Weise charakterisierte Transkriptionskontrollelement zeigt Homologie zur Ad Enhancer Stammsequenz und zur SV40 Enhancer Stammsequenz GGTGTGGAAAG (Weiher et al., 1983).

Die Ad5 Ela Region von 311-454 (-188 bis -45) vermittelt die konstitutive Expression eines normalerweise induzierbaren Gens (AdE2), was einer 10-15 fachen Steigerung entspricht. Dies zeigen Transfektionsansätze mit Plasmidkonstrukten, welche das Ad5 E2 Gen und besagte Ad5 Ela Sequenz enthalten (Imperiale et al., 1983).

Bei Ad12 sind die Enhancereigenschaften der Ela 5'-Region nicht so detailliert untersucht. Fest steht allerdings, daß die Sequenz bis zur AvaII Stelle (Pos. 525), vor das CAT Gen (Chloramphenicol Acetyltransferase) in den pSV0 Vektor (Gorman, 1982) kloniert (pAd12-la CAT; Kruczek & Dörfler, 1983), die Expression in gleichem Maß ermöglicht wie der frühe SV40 Promotor in pSV2 CAT, der ja den Enhancer enthält.

Der Ad12 Ela Promotor enthält 2 TATA Boxen ("Goldberg-Hogness Boxen"; Übersichtsartikel: Breathnach & Chambon, 1981) an den Positionen 276 und 414, Translationsstart ist bei ATG 502. Die Pos. 445 entspricht der cap-Stelle von Ad2 bzw. Ad5.

Die Ad12 Ela 5'-Region enthält die Ad Enhancer Stammsequenzen:

CGGAAGTGAAA	Pos. 248	bzw. -197	(bezogen auf Pos. 445)
CGGATGTGACG	153	-292	
AGGAAGTGGGG	119	-326	

Die Ad12 Region 148-177 entspricht, aufgrund der Homologie, dem Enhancer Element der Ad2 Ela 5'-Region (Pos. 149-178).

Zwei der drei Ad Enhancer Stammsequenzen liegen in der ITR. Dort, in einem GC-Cluster von Pos. 126-155 liegen auch 2 HpaII Stellen (Pos. 143 und 152). Werden diese methyliert, ist die CAT Expression des pAd12-la CAT Plasmids stark herabgesetzt

(Kruczek & Dörfler, 1983). Der gleiche Effekt wird durch die Methylierung der 7 HhaI Stellen im 5'-untranslatierten Bereich erzielt.

Der Adl2 DNA an der Integrationsstelle p9 fehlen demnach wichtige regulatorische Elemente. Kann es sein, daß die benachbarte Hamster DNA diese Funktionen übernommen hat?

Der Hamster DNA Anteil von p9 enthält ein Stück eines 4.5S RNA Gens und ein Transkriptionskontrollelement:

Die in p9 enthaltene Hamster DNA (Anhang 7.1.4.) hat eine ungewöhnliche Basenzusammensetzung, was möglicherweise die Voraussetzung für die Integration der Virus DNA war.

Sie ist Teil einer repetitiven, 4 kb langen Sequenz, die eines der zahlreichen Gene der für Hamster, Maus und Ratte spezifischen 4.5S RNA (Harada & Kato, 1980; Haynes et al., 1981) enthält. Diese 4.5S RNA, ein RNA-Polymerase III Produkt, zeigt Homologie zu den kurzen, hochrepetitiven Sequenzen der meisten Eukaryoten wie z.B. der menschlichen "Alu-Familie" (Houck et al., 1979) und kann mit polyadenylierter RNA assoziieren. p9 enthält das 3'-Ende eines 4.5S RNA Gens als bp 1 bis 39.

Der größere Teil der Hamster DNA von p9 wurde als Hamster Transkriptionskontrollelement (HTC) charakterisiert (Kruczek, Domrath & Garbrecht, zur Veröffentlichung eingereicht), denn ein Konstrukt aus der p9 Region 60-529 und der Adl2 Region 175-400 (PvuII Stelle) bzw. 175-525 (AvaII Stelle) in pSV0 zeigt gleich hohe CAT Expression wie das Konstrukt mit der vollständigen Adl2 Ela 5'-Region. Entfernt man dagegen aus letzterem die Adl2 Sequenz bis zur HpaII Stelle 152, ist die CAT Expression um mindestens 80% reduziert.

Es ist deshalb naheliegend, anzunehmen, daß dieses Hamster Transkriptionskontrollelement auch in der CLAC1 Tumorzelllinie die Expression des benachbarten Adl2 Ela Gens stimuliert. Dies

könnte ein entscheidender Faktor bei der Transformation der Hamsterzelle gewesen sein.

1.4.

Aufgabenstellung

TC7, eine epitheloide Nierenzelllinie von der Affenart *Cercopithecus aethiops* (Robb & Huebner, 1973), ist permissiv für das Simian Virus 40. Erstes Ziel der Arbeit war es, davon eine Genbank anzulegen, um daraus praktisch jeden beliebigen Genomabschnitt isolieren zu können.

Die Regulation der c-myc Genexpression ist noch weitgehend unverstanden. Das c-myc Gen des *Cercopithecus aethiops* sollte deshalb mit dem Ziel isoliert werden, den 5'-Bereich des Gens, einschließlich des 1. Exons, zu sequenzieren, um - im Vergleich mit den bereits bekannten c-myc Genen - Hinweise auf für die Expression wichtige DNA Sequenzen zu erhalten, denn diese zeichnen sich oft durch besonders gute Konservierung aus.

Desweiteren sollte die TC7 Genbank mit p9 als Probe durchgemustert werden, um im Genom des Affen Bereiche zu finden und zu charakterisieren, die zu dieser Hamster DNA Sequenz zwar nur entfernt verwandt sind, die aber eine ähnlich ungewöhnliche Basenzusammensetzung besitzen und möglicherweise gleiche Funktionen tragen, beispielsweise ein Transkriptionskontrollelement.

2.

MATERIAL

2.1.

Zellen, Phagen, Plasmide

TC7: Epitheloide Nierenzelllinie von der Affenart *Cercopithecus aethiops*, afrikanische grüne Meerkatze, African green monkey (Robb & Huebner, 1973). Subklon der Zelllinie CV1 (Jensen et al., 1964). Permissiv für das Simian Virus 40 (SV40).

E. coli BHB2688: N205 recA (λ imm434, cIts, b2, red3, Eam4, Sam7)/ λ

BHB2690: N205 recA (λ imm434 cIts, b2, red3, Dam15, Sam7)/ λ
(Hohn, 1979)

HB101: F⁻, hsdS20(r⁻_s, m⁻_s), supE44, ara-14, galK-2, lacY1, proA2, rspL20(str^R), xyl-5, mtl-1, λ^- , recA13
(Boyer & Roulland-Dussoix, 1969)

JM103: endA, Δ [lac, pro], thi-1, strA, sbcB15, hsdR4, supE, λ^- , [F' traD36, proA⁺B⁺, lacI^q Δ M15]
(Messing et al., 1981)

Q358: hsd(r⁻_s, m⁺_s), supF, $\emptyset$ 80^R

Q359: hsd(r⁻_s, m⁺_s), supF, $\emptyset$ 80^R, P2
(Karn et al., 1980)

Lambda 1059: h λ sBam1^o bl89 <int 29 ninL44 cI857 pac129>
 Δ [int-cIII] KH54, sR1 4^o, nin5, chi3
(Karn et al., 1980)

M13mpl0, M13mpl1 (Messing & Vieira, 1982)

M13mpl9 (Norrander et al., 1983)

pAT153 (Twigg & Sherratt, 1980)

pUC12 (Vieira & Messing, 1982)

2.2.

Enzyme, Chemikalien, Materialien, Geräte

Otto Aldag, Hamburg:
NZ-Amin

Amersham Buchler, Braunschweig:
[α - 32 P]dCTP (3000 mCi/mMol); [α - 32 P]dGTP (3000 mCi/mMol);
(α -[35 S]Thio)dATP (650 Ci/mMol); Hybond-N Nylon Membranfilter

J.T. Baker Chemical Co., Philipsburg, NJ, USA:
CsCl, Glycerol

Conrad Benda, Wiesloch:
UV Strahler N90, 254nm

Berthold, Wildbad:
Handmonitor LB 1210 B

Biochrom, Berlin:
"Seromed" Dulbeccos Medium

Biolabs, Schwalbach:
DNA Polymerase I; DNA Polymerase I großes (Klenow) Fragment;
Restriktionsendonukleasen; T4 DNA Ligase

Boehringer, Mannheim:
ATP; Chloramphenicol; Desoxynukleosidtriphosphate (dATP, dCTP, dGTP, dTTP); DNase I; Lysozym; alkalische Phosphatase;
Restriktionsendonukleasen; RNase A

BRL, Bethesda Research Laboratories, Eggenstein:
Restriktionsendonukleasen; Vanadyl Ribonukleosid Komplex

Calbiochem GmbH, Frankfurt/Main:
Dithiothreitol (DTT)

Collaborative Research, Inc., Waltham, Mass., USA:
Ologo(dT)-Cellulose Typ 2

Desaga, Heidelberg:
MinUVIS UV 254/366nm

Difco Laboratories, Detroit, Michigan, USA:
Bacto Agar; Bacto Trypton; Casamino Acids; Yeast Extract

DuPont, Bad Nauheim:
Röntgenfilm Exponierkassetten mit Quanta III Verstärkerfolien;
Ultrazentrifugen Sorvall OTD2 oder OTD-65B, Rotoren TV865 und
AH627, Ultrazentrifugenröhrchen; Laborzentrifugen Sorvall SS-3,
RC-5 oder RC-5C, Rotoren SS34, GSA und GS3

Eastman Kodak Company, Rochester N.Y., USA:
Röntgenfilme X-Omat S und X-Omat AR XAR5

Eppendorf Gerätebau, Netheler & Hinz GmbH, Hamburg:
Reaktionsgefäße 3810, 1.5ml, "Eppendorfhütchen"; Tischzen-
trifuge 5412, 10000 rpm, für Eppendorfhütchen

Eurolab, München:
Mikrofuge Röhrchen 0.5ml (snap cap)

Fuji Photo Film Co., Ltd., Japan:
Röntgenfilme RX

Falcon, Newtown, Connecticut, USA:
Blue Max 50ml Polypropylenröhrchen, steril

Flow Laboratories, Bonn:
fötales Kälberserum

Fluka AG, Feinchemikalien, Neu-Ulm:
Formamid puriss. p.a.

Gelman, Ann Arbor, Michigan, USA:
Acrodisc eingefaßte Membranfilter, 0.2 und 0.45µm
Porendurchmesser

Greiner, Nürtingen:
sterile 5 und 10ml Plastikröhrchen mit Deckel

Hereus Christ, Osterode:
Labofuge II

Hoechst, Frankfurt/Main:
Penicillin G Natrium; Streptomycinsulfat

LKB, Gräfelfing:
Makrophor Elektrophoresesystem mit Macrodrive 5 Spannungsquelle

E. Merck, Darmstadt:
Ammoniumperoxodisulfat; Borsäure; Bromphenolblau; n-Butanol;
CaCl₂; CHCl₃; Dimethylsulfoxid (DMSO); EDTA; HOAc; Ethidium-
bromid; Formaldehyd; Glucose; HCl; Na₂B₄O₇; NaCl; Na₂HPO₄;
NaH₂PO₄; Natriumpyrophosphat; NaOAc; NaOH; NH₄OAc; (NH₄)₂SO₄;
Isoamylalkohol; Isopropanol; KCl; KH₂PO₄; K₂HPO₄; KOAc; MgCl₂;
MgSO₄; Phenol; Polyethylenglycol 6000 (PEG 6000); Proteinase K
aus Pilzen; Saccharose; SDS (Dodecylhydrogensulfat Natrium-
salz); N,N,N',N'-Tetramethylethyldiamin (TEMED);
Thiaminiumdichlorid; Trinatriumcitrat; Trichloressigsäure;
Triton X 100

Nunc GmbH, Wiesbaden:
Impfnadeln und -ösen; Petrischalen; Zellkulturflaschen und
-schalen

Pharmacia Fine Chemicals, Uppsala, Schweden:
2',3'-Didesoxyribonukleosidtriphosphate (ddATP, ddCTP, ddGTP,
ddTTP); Ficoll 400; Sephadex G50 fine

Schleicher & Schüll GmbH, Dassel:
DEAE Cellulose Membranfilter NA45; Nitrocellulose Membranfilter BA85

Searle GmbH, Heusenstamm:
Flüssigkeitsszintillationszähler Modell 6892

Serva, Heidelberg:
Acrylamid; Gelatine; Glaswatte; Harnstoff; N-Laurylsarkosin Natriumsalz, 35%; 2-Mercaptoethanol; N,N'-Methylenbisacrylamid; Polyvinylpyrrolidon 90; RNA aus Hefe, niedermolekular; Siliconlösung; Visking Dialysesschläuche; Xylencyanol

Sigma Zentrifugenbau, Osterode:
Sigma 2MK Laborzentrifuge mit Rotor für 24 Eppendorfhütchen

Sigma Chemie, München:
Agarose; Ampicillin; 5-Brom-4-chlor-3-indolyl- β -D-galactopyranosid (Xgal); Isopropyl- β -D-thiogalaktosid (IPTG); gamma-(Methacryloxy)-propyltrimethoxysilan; Morpholinopropansulfonsäure (MOPS); Putrescin x 2HCl; Rinderserumalbumin (BSA), Spermidin x 3HCl; Tris

Whatman Ltd., Springfield Mill, Maidstone, Kentucky, USA:
Glasfilter GF/C, 3MM Filterpapier

Wifug, Stockholm, Schweden:
Tischzentrifuge Doctor

Carl Zeiss, Oberkochen:
Refraktometer 120599

Zinsser, Frankfurt:
Kunststoff Zählfläschchen; Quickscint 501

2.3.

Medien, Lösungen, Puffer**50x Denhardts**

Je 10g/L Ficoll 400, Polyvinylpyrrolidon und Rinderserumalbumin

LB

Medium: 10g Bacto Trypton, 5g Hefeextrakt, 5g NaCl ad 1L; autoklavieren

Platten: 15g/L Bacto Agar in LB Medium; autoklavieren; auf 55°C abkühlen, gewünschtes Antibiotikum (50µg/ml Ampicillin) zugeben; in Petrischalen gießen

Minimal Agarplatten

15g Bacto Agar, 10.5g K_2HPO_4 , 4.5g KH_2PO_4 , 1g $(NH_4)SO_4$, 0.5g Natriumcitrat x $2H_2O$ und 0.2g $MgSO_4$ x $7H_2O$ ad 990ml; autoklavieren; 5µg Thiamin x $2HCl$ und 2g Glucose ad 10ml sterilfiltriert zugeben; in Petrischalen gießen

NZCY

Medium: 10g NZ-Amin, 4.5g Hefeextrakt, 1g Casamino-säuren, 5g NaCl und 2g $MgSO_4$ x $7H_2O$ ad 1L

Platten: 10g/L Bacto Agar in NZCY Medium; autoklavieren; in Petrischalen gießen

Weichagarose: 6g/L Agarose (Sigma Typ I) in NZCY Medium; autoklavieren

SM

0.1M NaCl, 10mM $MgSO_4$, 50mM Tris pH 7.5, 0.01% Gelatine; autoklavieren

20x SSC

3M NaCl, 0.3M Trinatriumcitrat

20x SSPE

3.6M NaCl, 0.2M $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$, 20mM EDTA; pH 7.4 einstellen durch Zugabe von 10N NaOH; autoklavieren

TE

10mM Tris pH 8.0, 1mM EDTA

Weitere Angaben gelten entsprechend der Definition von TSE.

10x TEB

0.89M Tris, 25mM EDTA, 0.89M Borsäure

Tris Saline

137mM NaCl, 5mM KCl, 6mM Na_2PO_4 , 250mM Tris; durch Zugabe von 1N HCl pH 7.4 einstellen; mit 0.3g/L Penicillin und 0.1g/L Streptomycin versetzen; sterilfiltrieren

TSE

T steht für Tris, S für NaCl und E für EDTA; dahinter sind die Konzentrationen (mM) der Komponenten und der pH Wert der Tris Stammlösung angegeben.

Beispiel: TSE 10/100/5 pH 7.5 = 10mM Tris pH 7.5, 100mM NaCl, 5mM EDTA

YT

Medium: 8g Bacto Trypton, 5g Hefeextrakt und 5g NaCl ad 1L; autoklavieren

Platten: 15g/L Bacto Agar in YT Medium; autoklavieren; in Petrischalen gießen

Weichagar: 6g/L Bacto Agar in YT Medium; autoklavieren

3.

METHODEN

Falls nichts anderes angegeben, handelt es sich um Methoden wie sie in Maniatis et al. (1982) oder in verschiedenen Bänden der Reihe "Methods in Enzymologie" (Academic Press, Inc. New York, London) ausführlich beschrieben sind.

3.1.

Zellen, Bakterien, Phagen

3.1.1.

Züchtung von TC7 Zellen

TC7 Zellen wurden als Einschichtkulturen in 200ml Zellkulturflaschen (75cm²) oder in 24 x 24cm Zellkulturschalen in 40 bzw. 400ml Dulbeccos Medium (Dulbecco & Freeman, 1959) mit 5% fötalem Kälberserum gezüchtet.

3.1.2.

Züchtung von Bakterien

Bakterien wurden, ausgehend von Einzelkolonien, in Flüssigkultur oder auf Agarplatten in LB-, Minimal-, NZCY- oder YT Medium gezüchtet. Reine Stämme wurden für kurze Zeit auf Agarplatten bei 4°C aufbewahrt oder als mit 20% Glycerol versetzte Flüssigkulturen (Bakterienstöcke) bei -80°C eingefroren.

3.1.3.

Züchtung von Phagen

(Züchtung von M13 Phagen: siehe 3.3.6.)

3.1.3.1.

Vermehrung von Lambda Phagen auf Agarplatten

Eine Übernachtskultur (NZCY Medium) der Wirtsbakterien (Q358 oder Q359) abzentrifugieren. Das Pellet in 1/5 Volumen 10mM MgSO_4 suspendieren. Eine Konzentration von ca. 1.5×10^9 Zellen/ml einstellen ($\text{OD}_{600} = 2$). Die Bakteriensuspension ist, bei 4°C aufbewahrt, eine Woche lang verwendbar.

100µl Bakteriensuspension und 5-50µl Phagenlösung vermischen und 15 min bei 37°C schütteln. 2.5ml flüssige, 47-50°C warme NZCY Weichagarose dazumischen und sofort auf 37°C warme NZCY Agarplatten ausplattieren. Über Nacht oder mindestens 12 h bei 37°C inkubieren.

Die Mengenangaben beziehen sich auf die Verwendung von Petrischalen mit 9.5cm Durchmesser und sind gegebenenfalls auf andere Größen adaptierbar.

3.1.3.2.

Titration

Von einer Phagenlösung eine Verdünnungsreihe in SM machen. Je 5µl der in Frage kommenden Verdünnungsstufen zur Vermehrung einsetzen. Plaques zählen und den Titer berechnen.

3.1.3.3.

Lambda Phagenstock anlegen

10^5 pfu wie beschrieben vermehren. Auf den so gut wie vollständig lysierten Zellrasen 5ml SM geben. Mindestens 4h bei 4°C stehen lassen. Danach die Flüssigkeit in ein verschließbares Röhrchen überführen und mit $100\mu\text{l}$ CHCl_3 versetzen.

Ein Lambda Phagenstock enthält 10^9 - 10^{10} pfu/ml und ist bei 4°C jahrelang stabil.

3.1.3.4.

Plaque Reinigung

Einen einzelnen, isolierten Plaque mitsamt dem darunterliegenden Agar in 0.25ml SM einbringen. Mindestens 4 h bei 4°C stehen lassen. Von $50\mu\text{l}$ dieser Lösung (ca. 10^5 pfu) einen Phagenstock anlegen.

3.2.

Grundtechniken

3.2.1.

Phenolextraktion von Nukleinsäure Lösungen

Phenol mit TSE (10mM Tris pH 7.4-8.0, 1 - 5mM EDTA, 0 - 100mM NaCl) oder 50mM $\text{Na}_2\text{B}_4\text{O}_7$ sättigen.

Die Nukleinsäure Lösung nacheinander mit je einem Volumen gesättigtem Phenol, Phenol/ CHCl_3 (1:1) und CHCl_3 /Isoamylalkohol (24:1) extrahieren. Diese Extraktion wird im folgenden "phenolisieren" genannt.

Bei kleinen Volumina empfiehlt es sich, auch das Phenol/ CHCl_3 mit Puffer zu sättigen.

3.2.2.

Fällen von Nukleinsäuren

Wenn im Text nichts anderes angegeben, bedeutet "fällen" oder "repräzipitieren": 0.1 Volumen 3M NaOAc pH 5.2 und 2.5 Volumen EtOH zur Nukleinsäure Lösung geben, über Nacht bei -20°C oder 1 h (bei kleinem Volumen genügen 10 min) bei -80°C halten. 20 min bei 4°C und 10000 rpm im SS34 Rotor oder in einer Eppendorf Zentrifuge pelletieren. Den Überstand abgießen oder abpipettieren.

Für die Fällung kleiner DNA Fragmente empfiehlt es sich, zusätzlich 10mM MgCl_2 einzustellen.

3.2.3.

Konzentrationsbestimmung von Nukleinsäure Lösungen

Die Absorption einer Probe bei 260 und 280nm Wellenlänge messen. 1 A_{260} entspricht 50 $\mu\text{g/ml}$ dsDNA, 40 $\mu\text{g/ml}$ ssDNA bzw. RNA oder 20 $\mu\text{g/ml}$ Oligonukleotide.

Reine Nukleinsäure Präparationen haben einen A_{260}/A_{280} Quotienten von 1.8.

3.2.4.

Spaltung von DNA durch Restriktionsendonukleasen

In erster Linie gelten die vom Hersteller empfohlenen Bedingungen. Diese werden durch Zugabe 5-fach konzentrierter Stammlösungen eingestellt.

Ist eine DNA Lösung mit RNA verunreinigt, setzt man dem Restriktionsansatz 20 $\mu\text{g/ml}$ DNase freie RNase A zu. Dies verbessert die Spaltung und verhindert das Überdecken von kleinen DNA Fragmenten nach der Auftrennung mittels Gelelektrophorese.

Läßt sich eine DNA wider erwarten nicht spalten, so hilft in den meisten Fällen eine Dialyse gegen TE.

Spaltet man mit 2 Restriktionsenzymen, kann man dies gleichzeitig tun, wenn die optimalen Bedingungen ähnlich sind. Ist dies nicht der Fall, spaltet man zuerst mit dem Enzym, das die niedrigere Salzkonzentration verlangt und stellt danach die Bedingungen für das zweite Enzym ein.

Wenn eines der beiden Enzyme KCl benötigt, und man kann, wie etwa bei der doppelten Spaltung eines Polylinkers (Bereich eines Vektors, der Erkennungsstellen für mehrere, verschiedene Restriktionsendonukleasen enthält), nicht unterscheiden, ob eines oder beide Enzyme geschnitten haben, empfiehlt es sich, die DNA zwischendurch zu fällen und zu waschen.

Wird nach der Inkubation 0.1 Volumen 0.5M EDTA zugesetzt, ist ein Aufbewahren des Ansatzes bei 4°C über längere Zeit hinweg möglich.

3.2.5.

Gelelektrophorese von Nukleinsäuren

3.2.5.1.

DNA

DNA Moleküle der Größenordnung von 0.2 bis über 20kb wurden durch Elektrophorese in (meist horizontal angeordneten) Agarosegelen charakterisiert bzw. aufgetrennt.

Dazu die DNA Probe mit 0.1 Volumen 60% Saccharose, 0.4% Bromphenolblau versetzen, auf ein Gel auftragen und durch Anlegen eines elektrischen Feldes von maximal 5V/cm entsprechend lange der Elektrophorese unterwerfen. Das Gel besteht, je nach Größenverteilung der DNA Moleküle, aus 0.6 bis zu 1.8% Agarose in 40mM Tris pH7.8, 5mM NaOAc, 1mM EDTA, 0.5µg/ml Ethidiumbromid).

Die DNA wird durch Bestrahlung mit UV Licht von 254nm sichtbar gemacht und photographiert. Soll die DNA weiter verwendet werden, wird sie bei UV Licht von 366nm betrachtet und aus dem Gel isoliert (3.2.6.).

DNA Moleküle von 20 bis 600 bp Länge wurden durch Elektrophorese in vertikal angeordneten Polyacrylamid Gelen charakterisiert.

Dazu die Probe mit 0.1 Volumen 60% Saccharose, 0.4% Bromphenolblau, 0.4% Xylencyanol versetzen, auf ein Gel auftragen und durch Anlegen eines elektrischen Feldes von ca. 5V/cm entsprechend lange der Elektrophorese unterwerfen. Das Gel besteht aus 8 bis zu 12% Acrylamid/N,N'-Methylenbisacrylamid = 29:1, 89mM Tris, 2.5mM EDTA, 89mM Borsäure, polymerisiert durch 0.06% Ammoniumpersulfat und 0.03% TEMED.

Die DNA durch Baden in Ethidiumbromidlösung (ca. 1µg/ml) anfärben und im UV Licht (254nm) photographieren.

3.2.5.2.

RNA

Elektrophorese in Agarose/Formaldehyd Horizontalgelen: 1.3% Agarose in MOPS Puffer (40mM Morpholinopropansulfonsäure, 10mM NaOAc, 1mM EDTA; pH 7.0 eingestellt durch NaOH) mit 2.2M Formaldehyd.

Dazu die (gefällte) RNA in 32µl Probenpuffer (3µl 10x MOPS Puffer, 5.3µl Formaldehyd, 15µl Formamid, 3.2µl 50% Saccharose, 0.4% Bromphenolblau, 0.4% Xylencyanol, 5.5µl H₂O) aufnehmen, 20 min auf 55°C halten, auf das Gel auftragen und durch Anlegen eines elektrischen Feldes von ca. 5V/cm der Elektrophorese unterwerfen. Den Puffer der beiden Kammern ständig umpumpen. Für eine Spur des Gels 4µg polyadenylierte oder 40µg cytoplasmatische RNA einsetzen.

3.2.6.

Isolierung von DNA Fragmenten aus Agarose Gelen

Produktbegleitende Vorschrift des DEAE Cellulose Membran Herstellers (Schleicher & Schüll).

Bei UV Licht (366nm) vor der Bande des gewünschten DNA Fragments einen Schnitt in das Gel machen. Dabei die DNA möglichst wenig der Strahlung aussetzen. Ein Stückchen DEAE Cellulose Membran in den Schnitt einsetzen. Die Elektrophorese fortsetzen, bis die Bindung der DNA an die Membran vollständig ist. Die Membran aus dem Gel ziehen und durch Schütteln in TSE 20/150/0.1 pH 8.0 von Agaroseresten befreien. Die Membran in einem Mikrofuge Röhrchen in 200µl TSE 20/1000/0.1 pH 8.0 eintauchen und 30 min bei 60°C halten. Die Lösung in ein Eppendorfhütchen überführen. Diesen Elutionsschritt mit 50µl wiederholen. Die so erhaltene DNA Lösung 2x mit n-Butanol extrahieren und ohne weitere Zugabe von NaOAc fällen. Die DNA in TE lösen, repräzipitieren und wieder in TE aufnehmen. Bei -20°C aufbewahren.

Ausbeute: ca. 90%.

3.3.

Isolierung von Nukleinsäuren

3.3.1.

Isolierung von hochmolekularer DNA aus Gewebekultur Zellen

Zellen in Schalen bis zur Konfluenz (2×10^5 Zellen/cm²) wachsen lassen. 3x mit eiskalter Tris Saline waschen. Zellen mit einem Plastikwischer abschaben, in Zentrifugenröhrchen überführen, 10 min bei 4°C und 5700 rpm in einer Tischzentrifuge sedimentieren. Niederschlag pro 10^8 Zellen in 1ml Tris Saline suspendieren. 4 Volumen 125mM EDTA pH 8.0, 125 µg/ml Proteinase K,

6.25% N-Laurylsarkosin zugeben und 4 h bei 37°C inkubieren. Die Lösung phenolisieren und mit 2.5 Volumen EtOH versetzen. Die DNA an einem Glasstab sammeln (aufwickeln), über Nacht in 70% EtOH bei 4°C halten. Danach an der Luft trocknen lassen und schließlich in TE aufnehmen.

Konzentration bestimmen. Bei 4°C aufbewahren.

Ausbeute ca. 12mg/10⁸ Zellen.

3.3.2.

Isolierung von cytoplasmatischer RNA aus Gewebekultur Zellen

Zellen in Kulturschalen züchten. Medium abgießen. Zellschicht 3x mit eiskalter Tris Saline waschen, mit einem Plastikwischer abschaben, in ein Zentrifugenröhrchen überführen und 10 min bei 4°C und 5700 rpm in einer Tischzentrifuge sedimentieren. Das Pellet pro 10⁸ Zellen (ca. 500cm² Zellrasen) in 2.2ml Lysis Puffer (frisch zubereitet: 140mM NaCl, 1.5mM MgCl₂, 10mM Tris pH8.6, 0.5% Triton X 100, 10mM Vanadyl Ribonukleosid Komplex) suspendieren. Die Zellsuspension mit 1 Volumen Lysis Puffer, der zusätzlich 24% Saccharose und zusätzlich 0.5% Triton X 100 enthält, unterschichten und 20 min bei 4°C und 9500 rpm im SS34 Rotor zentrifugieren. Die obere Phase in ein frisches Röhrchen überführen, mit 1 Volumen 0.2M Tris pH 7.5, 0.3M NaCl, 25mM EDTA, 2% SDS versetzen, 200µg/ml Proteinase K einstellen, 30 min bei 37°C inkubieren, phenolisieren und fällen. Das Pellet in H₂O aufnehmen. Die RNA Konzentration der Lösung bestimmen. Repräzipitiert als Suspension bei -20°C aufbewahren.

3.3.3.

Isolierung polyadenylierter RNA

0.4g Oligo(dT)-Cellulose in Puffer A (20mM Tris pH 7.6, 0.5M

NaCl, 1mM EDTA, 0.1% SDS) suspendieren und damit eine 1ml Säule beladen. Diese mit 4ml H₂O, mit 4ml 0.1N NaOH, 5mM EDTA, und wieder mit H₂O waschen, bis der Durchlauf einen neutralen pH Wert hat. Schließlich mit 6ml Puffer A waschen.

2mg cytoplasmatische RNA, gelöst in ca. 0.4ml H₂O, 5 min bei 65°C halten. 1 Volumen 2x Puffer A dazugeben, auf RT kühlen und auf die Säule laden. Den Durchlauf wieder auf 65°C erhitzen, kühlen und auftragen. Die Säule mit 8ml Puffer A und 5ml Puffer B (20mM Tris pH 7.6, 0.1M NaCl, 1mM EDTA, 0.1% SDS) waschen. Die Elution der nicht polyadenylierten RNA kann durch Konzentrationsmessung der einzelnen, durchgetropften Säulenvolumen verfolgt werden.

Die polyadenylierte RNA mit 6ml Puffer E (10mM Tris pH 7.5, 1mM EDTA, 0.05% SDS) eluieren. Konzentration bestimmen. Die polyadenylierte RNA, gefällt als Suspension, bei -20°C aufbewahren. Ausbeute: 2% der eingesetzten RNA.

3.3.4.

Isolierung von Plasmid DNA

3.3.4.1.

CsCl Gleichgewichtszentrifugationsmethode für größere Mengen

Eine Übernachtskultur des gewünschten, Plasmid tragenden Stamms (Einzelkolonie oder Bakterienstock) in LB Medium mit dem entsprechenden Antibiotikum bei 37°C züchten. Mit 10ml davon die 1L Hauptkultur animpfen. Bei 37°C schütteln, bis eine OD₆₀₀ von 0.8 bis 1.0 erreicht ist. 160µg/ml Chloramphenicol, gelöst in EtOH, zugeben. Über Nacht weiter schütteln. 10 min bei 4°C und 7000 rpm im GS3 Rotor zentrifugieren. Überstand verwerfen. Pellet bei 0°C in 6ml 25% Saccharose, 50mM Tris pH 8.0 suspendieren, in Zentrifugenbecher überführen und 5mg Lysozym, gelöst in 1ml 0.25M Tris pH 8.0, zugeben. 5 min auf Eis stehen lassen,

während dessen gelegentlich schwenken. 1ml 0.25M EDTA pH 8.0 zugeben. Wieder 5 min auf Eis stehen lassen, gelegentlich schwenken. 8ml Lysis Mix (50mM Tris pH 8.0, 62.5mM EDTA pH 8.0, 0.2% Triton X 100) zugeben. Sobald die Lösung viskos und etwas durchscheinend wird, 30 min bei 4°C und 20000 rpm im SS34 Rotor zentrifugieren. Den klaren Überstand in ein Plastikröhrchen überführen, dabei das Volumen exakt bestimmen. Pro ml 0.95g CsCl darin lösen und 50µl 20mg/ml Ethidiumbromid zugeben. Mindestens 18 h bei 18°C und 47000 rpm im TV865 Rotor zentrifugieren. Die Plasmidbande mit einer Spritze ernten. Durch mehrere Extraktionen mit Isopropanol, gesättigt mit 5M NaCl, das Ethidiumbromid entfernen. Dialysieren gegen TE, phenolisieren, erneut gegen TE dialysieren. Die DNA Konzentration der Plasmid Lösung bestimmen. Bei 4°C aufbewahren. Die Ausbeute beträgt gewöhnlich 1.5-3mg.

3.3.4.2.

Schnellaufschluß Methode für kleine Mengen

1.5ml einer Übernachtskultur in ein Eppendorfhütchen pipettieren, 1 min zentrifugieren. Das Pellet in 100µl 0°C kalter Lysozymlösung (frisch zubereitet: 4µg/µl Lysozym, 50mM Glucose, 10mM EDTA, 25mM Tris pH 8.0) suspendieren, 5 min bei RT halten. Danach 200µl 0.2N NaOH, 1% SDS (frisch zubereitet) zugeben und 5 min bei 0°C halten.

Mit 150µl Acetat Puffer (60% 5M KOAc, 11.5% HOAc, 28.5% H₂O) versetzen und 5 min bei 0°C halten. 5 min zentrifugieren. Den Überstand phenolisieren. Bei RT 2 Volumen EtOH zugeben, 2 min stehen lassen und 5 min zentrifugieren. Das Pellet mit 1ml 70% EtOH waschen, kurz lyophilisieren und in 50µl TE, das 20µg/ml DNase freie RNase A enthält, aufnehmen. Die Plasmid Lösung bei -20°C aufbewahren.

3.3.5.

Isolierung von Lambda DNA

3.3.5.1.

CsCl Gleichgewichtszentrifugationsmethode für größere Mengen

Eine 50ml Übernachtskultur der Wirtsbakterien (Q358 oder Q359; NZCY Medium) abzentrifugieren. In 5ml 10mM MgSO₄ suspendieren. Die Konzentration der Zellen bestimmen ($1 \text{ OD}_{600} = 8 \times 10^8$ Zellen) und auf ca. 5×10^9 Zellen/ml einstellen.

10^{10} Zellen mit 5×10^8 Lambda Phagen gut vermischen, 15 min bei 37°C leicht schütteln, zu 500ml 37°C warmem NZCY Medium pipetieren und bei 37°C heftig schütteln, bis nach ca. 4 h Lysis eintritt. 10ml CHCl₃ und 60ml 5M NaCl zugeben, weitere 5 min schütteln und in Eiswasser kühlen. Den Ansatz - ohne das CHCl₃ - in Zentrifugenbecher überführen und 15 min bei 4°C, 5000 rpm im GSA Rotor zentrifugieren. Den Überstand bei RT unter Rühren mit 60g Polyethylenglykol (PEG 6000) versetzen, 90 min auf Eis stehen lassen und 20 min bei 4°C, 5000 rpm im GSA Rotor zentrifugieren. Das Pellet in 7ml 50mM Tris pH 7.5, 10mM MgSO₄ suspendieren, 1 Volumen CHCl₃ zugeben, 5min schütteln und 20 min bei 10000 rpm im SS34 Rotor zentrifugieren. Die wässrige Phase abnehmen, das Volumen exakt bestimmen, 0.75g CsCl/ml zugeben, Dichte im Refraktometer kontrollieren (Sollwert: 1.50g/ml) und gegebenenfalls einstellen, 2 h bei 20°C und 47000 rpm im TV865 Rotor zentrifugieren. Die Bande aus Phagenpartikeln mit einer Spritze ernten und gegen TSE 10/100/5 pH 7.5 dialysieren. Die Phagenlösung auf 100µg/ml Proteinase K und 0.5% SDS einstellen, 1 h bei 37°C inkubieren, phenolisieren, fällen, mit 70% EtOH waschen, kurz lyophilisieren, in TE 10/1 pH 7.5 aufnehmen. Die DNA Konzentration bestimmen. Ausbeute ca. 500µg Phagen DNA.

3.3.5.2.

Schnellpräparation für kleine Mengen

1.6x10⁸ Bakterien mit 3x10⁶ Lambda Phagen infizieren (siehe 3.3.5.1.), 4ml NZCY Medium zugeben und ca. 9 h (oder über Nacht) bei 37°C heftig schütteln. 100µl CHCl₃ zugeben und 15 min weiterschütteln. Den Ansatz 10 min bei 4°C und 8000 rpm im SS34 Rotor zentrifugieren. Den Überstand nehmen, je 1µg/ml RNase A und DNase I einstellen und 30 min bei 37°C inkubieren. 1 Volumen 20% PEG 6000, 2M NaCl in SM zugeben und 1 h in Eiswasser stehen lassen. 20 min bei 4°C und 10000 rpm im SS34 Rotor zentrifugieren. Den Überstand verwerfen. Die präzipitierten Phagen in 0.5ml SM suspendieren, in ein Eppendorfhütchen überführen und 2 min bei 4°C zentrifugieren. Den Überstand in ein frisches Eppendorfhütchen überführen, 5µl 10% SDS und 5µl 0.5M EDTA pH 8.0 zugeben. 15 min bei 68°C inkubieren, phenolisieren, 1 Volumen Isopropanol zugeben, 20 min bei -80°C halten, auftauen und 15 min bei 4°C zentrifugieren. Das Pellet mit 70% EtOH waschen, trocknen und in 50µl TE 10/1 pH 7.5 aufnehmen. Bei -20°C aufbewahren.

Ausbeute ca. 2µg.

3.3.6.

Isolierung von M13 DNA

3.3.6.1.

Schnellaufschluß Methode zur Isolierung von M13 ssDNA

Den ausgestochenen Plaque eines (rekombinanten) Phagens in 1ml YT Medium überführen, dazu 10µl einer exponentiell wachsenden JM103 Kultur geben. Über Nacht bei 37°C schütteln. Die Suspension in ein Eppendorfhütchen überführen und 5 min zentrifugieren.

ren. Den Überstand in ein frisches Eppendorfhütchen überführen. (Das Pellet und etwas verbleibender Überstand dient als Stock und wird bei 4°C aufbewahrt.) 200µl 20% PEG 6000, 2M NaCl zum Überstand geben, 15 min bei RT stehen lassen, 5 min zentrifugieren. Überstand verwerfen. Pellet in 100µl TSE 20/10/0.1 pH 7.5 suspendieren, phenolisieren und fällen. In 50µl TSE aufnehmen und erneut fällen. Schließlich in 25µl TSE aufnehmen und bei -20°C aufbewahren.
Ausbeute ca. 5µg.

3.3.6.2.

Isolierung von M13 RF DNA

7ml YT Medium mit 100µl einer exponentiell wachsenden Kultur von JM103 und 25µl eines M13 Phagenstocks animpfen, 4 h bei 37°C schütteln und 5 min bei 7000 rpm im SS34 Rotor abzentrifugieren. Mit 5ml dieses Überstands, 5ml exponentiell wachsende JM103 in 500ml YT Medium infizieren. Über Nacht bei 37°C schütteln. Die Kultur der infizierten JM103 Zellen wie für eine Plasmid Präparation (3.3.4.1.; halber Ansatz) aufarbeiten.
Ausbeute ca. 250µg.

Kleine Mengen M13 RF DNA können aus infizierten Bakterien wie ein Plasmid nach der Schnellaufschluß Methode (3.3.4.2.) isoliert werden.

3.4.

Vorbereitung der DNA für die Klonierung

3.4.1.

Präparation von Fragmenten zellulärer DNA eines bestimmten Größenbereichs

Ca. 200µg zelluläre DNA unvollständig mit einem passenden Restriktionsenzym (z.B. Sau3A oder MboI) schneiden. Dazu 5-6 Teilansätze bei konstanter Zeit mit verschiedenen Enzymkonzentrationen inkubieren, oder einem Ansatz mit bestimmter Enzymkonzentration nach verschiedenen Zeiten Anteile entnehmen und die Reaktion abstoppen. Die genauen Bedingungen dazu werden zuvor in Probeansätzen getestet.

Die Ansätze mit der gespaltenen DNA vereinigen, phenolisieren, fällen und in 200µl TE pH 8.0 aufnehmen. Die Lösung 20 min auf 68°C erhitzen, abkühlen und auf einen 38.5ml 10-40% Saccharose Gradienten in 1M NaCl, 20mM Tris pH 8.0, 10mM EDTA schichten. Den Gradienten 24 h bei 20°C und 27000 rpm in einem Ausschwingrotor (DuPont AH627 oder Beckman SW27) zentrifugieren und danach in 1ml Fraktionen tropfen. Von jeder Fraktion 25µl der Gelelektrophorese unterwerfen und Menge sowie Größenbereich der DNA durch Vergleich mit einer genau definierten DNA Probe ("Marker") bestimmen. Passende Fraktionen vereinigen, dialysieren (z.B. gegen TSE 10/5/1 pH 7.5), fällen und in einer Konzentration von ca. 0.3µg/µl aufnehmen.

3.4.2.

Präparation von Lambda Vektorarmen mit BamHI Enden

Ca. 150µg Lambda 1059 DNA in 150-300µl 100mM NaCl, 10mM Tris pH 7.5, 10mM MgCl₂ 1 h bei 42°C inkubieren, um zu erreichen, daß

möglichst viel DNA in langen Konkatemeren vorliegt. Danach mit BamHI spalten, phenolisieren, fällen und in 400µl TSE 10/100/1 pH 8.0 wieder aufnehmen. Je 200µl auf einen Gradienten aufbringen. Der Gradient und das weitere Vorgehen ist unter 3.4.1. beschrieben.

3.4.3.

Präparation von Plasmid- oder M13 Vektoren

10µg Plasmid DNA oder M13mp.. RF DNA mit einem oder zwei Restriktionsenzymen spalten. Danach den Ansatz so gut wie möglich auf Phosphatase Bedingungen (Herstellerangaben) einstellen und mit 10 Enzymeinheiten alkalischer Phosphatase inkubieren. Anschließend 15 min auf 65°C halten. Das Volumen auf 200µl einstellen. Den Ansatz phenolisieren, fällen, mit 70% EtOH waschen, kurz lyophilisieren und in TE aufnehmen, daß die DNA Konzentration 25µg/µl beträgt. Bei -20°C aufbewahren. Falls ein Transformationstest mit religiertem Vektor allein zu viele resistente Bakterienkolonien ergibt, empfiehlt es sich, die Phosphatasebehandlung - diesmal unter den exakten Bedingungen - zu wiederholen.

3.5.

Klonierungsschritte

3.5.1.

Ligase Reaktionen

3.5.1.1.

Verknüpfung von Lambda Phagenarmen mit Fragmenten zellulärer DNA

6.7µl = 2.3µg gereinigte Lambda 1059 Phagenarme

1.0µl 100mM MgCl₂

1.0µl 100mM Tris pH 7.4

1.0µl 20mM ATP

1.0µl T4 DNA Ligase = 400 Enzymeinheiten

4.0µl = 1.3µg zelluläre DNA, 12-20kb Fragmente

H₂O ad 20µl

24 h bei 16°C inkubieren

3.5.1.2.

Verknüpfung von Plasmid- oder M13 Vektoren mit der zu klonierenden DNA

50ng Vektor und die entsprechend molare Menge eines DNA Fragments oder Fragmentgemisches in 20µl mit T4 DNA Ligase inkubieren. Bedingungen gemäß Herstellerangaben.

3.5.2.

Transformation von Bakterienzellen

3.5.2.1.

Präparation transformationskompetenter Bakterien (*E.coli* HB101)

200ml LB Medium mit 2ml einer Übernachtskultur von *E.coli* HB101 animpfen. Bei 37°C schütteln, bis eine OD₆₀₀ von 0.5 erreicht ist. Die Bakterienkultur in einem Kochsalz/Eis Bad (-15°C) auf 0-5°C abkühlen und 10 min bei 4°C und 8000 rpm im GSA Rotor zentrifugieren. Das Pellet vorsichtig in 50ml eiskaltem 0.1M MgCl₂ suspendieren. 10 min bei 4°C und 2500 rpm im SS34 Rotor zentrifugieren. Das Pellet in 50ml eiskaltem 0.1M CaCl₂ suspendieren. 20 min auf Eis stehen lassen. Erneut 10 min bei 4°C und 2500 rpm zentrifugieren. Die Bakterienmasse in 10 ml 20% Glycerol, 80mM CaCl₂ suspendieren. Portionen von 100µl in flüssigem Stickstoff einfrieren und bei -80°C aufbewahren. Transformationsfrequenz ca. $1.2 \times 10^6 / \mu\text{g}$ pUC12 DNA.

3.5.2.2.

Transformation von *E.coli* HB101

100µl kompetente Bakterien auf Eis auftauen. Mit 5-50ng Plasmid DNA eines Ligase Ansatzes in 10-20µl Ligase Puffer (oder TE) vermischen. 1 h auf Eis stehen lassen. 2 min bei 42°C halten. Den Ansatz, verteilt auf bis zu 10 LB Platten mit dem entsprechenden Antibiotikum (für pUC Plasmide 50µg/ml Ampicillin) ausplattieren. Über Nacht bei 37°C inkubieren.

3.5.2.3.

Transformation von *E.coli* JM103 durch M13 RF DNA

E.coli JM103 auf Minimal Agar ausstreichen. Bei 37°C inkubieren. (Es dauert zwei Tage bis Kolonien zu sehen sind.) 5ml YT Medium mit einer Einzelkolonie animpfen und über Nacht bei 37°C schütteln. 1ml dieser Übernachtskultur zu 50ml YT Medium geben. Bei 37°C schütteln, bis eine OD₆₀₀ von 0.3-0.4 erreicht ist. 5ml YT Medium (mit einer Spur davon) frisch animpfen. Die 50ml Kultur 5 min bei 7000 rpm im SS34 Rotor zentrifugieren. Die sedimentierten Bakterien in 24ml eiskaltem 50mM CaCl₂ suspendieren, 20 min auf Eis stehen lassen und 5 min bei 4°C und 7000 rpm im SS34 Rotor zentrifugieren. Das Pellet in 5ml eiskaltem 50mM CaCl₂ suspendieren. Diese Gesamtmenge kompetenter Zellen reicht für 16 Transformationsansätze.

0.3ml Bakteriensuspension mit 10ng M13 RF DNA eines Ligase Ansatzes (in ca. 5µl Ligase Puffer oder TE) vermischen, 40 min auf Eis stehen lassen und 2 min bei 42°C halten. 10µl 0.1M IPTG, 50µl 2% Xgal in Dimethylformamid, 200µl (der inzwischen gewachsenen) JM103 Kultur und 3ml 45°C warmen YT Weichagar zugeben. Sofort auf eine vorgewärmte YT Platte gießen (gleichmäßig verteilen). Über Nacht bei 37°C inkubieren. Rekombinante Phagen (weiße Plaques) vermehren (3.3.6.).

3.5.3.

In vitro Verpackung von Lambda DNA

Mit je einer einzelnen, getesteten (Wachstum bei 32°C, Lyse bei 42°C; wegen des temperatursensitiven Repressors) Kolonie der *E.coli* Stämme BHB2688 und BHB2690 5ml LB Medium animpfen und über Nacht bei 32°C schütteln.

Damit je 200ml vorgewärmtes LB Medium zu OD₆₀₀ = 0.5 inokulieren und bei 32°C schütteln, bis eine OD₆₀₀ = 0.3 erreicht ist.

Die beiden Kulturen in einem 90°C heißen Wasserbad auf 44-45°C erwärmen. Dabei heftig schwenken und die Temperatur des Mediums durch Eintauchen eines Thermometers messen. Die beiden Kulturen 25 min bei 45°C und anschließend 3 h bei 37°C schütteln. Danach die Induktion der Phagengene testen: kleine Proben der beiden Kulturen werden nach Zugabe eines Tropfens CHCl_3 in wenigen min klar. Die beiden Kulturen (positive Induktionstests vorausgesetzt) in Eiswasser kühlen und vereinigen. 10 min bei 4°C und 5000 rpm im GSA Rotor zentrifugieren. Danach im Kühlraum (4°C) weiterarbeiten. Den Überstand möglichst vollständig entfernen. Die Hälfte der Zellen in 380 μl Verpackungspuffer (40mM Tris pH 8.0, 10mM Spermidin, 10mM Putrescin, 0.1% β -Mercaptoethanol, 7% Dimethylsulfoxid) suspendieren, 5 μl 0.1M ATP zugeben, 20 μl -weise in Eppendorfhütchen abfüllen und in flüssigen Stickstoff eintauchen (Verpackungsmix 1). Die andere Hälfte der Zellen ebenfalls in Verpackungspuffer suspendieren, 5 μl 0.1M ATP, 5 μl 1mg/ml DNase I und 5 μl 1M MgCl_2 zugeben, auf 6 Eppendorfhütchen verteilen und in flüssigen Stickstoff tauchen (Verpackungsmix 2).

Die Mixe bei -80°C aufbewahren. Dadurch ist die volle Verpackungsaktivität für mindestens 1 Jahr gewährleistet.

20 μl Verpackungsmix 1, während des Auftauens auf Eis, mit bis zu 1 μg Lambda DNA (bzw. rekombinanter Lambda DNA aus einem Ligase Ansatz) in maximal 5 μl Lösung vermischen. 40 min bei 37°C inkubieren. 20 μl Verpackungsmix 2, kurz zuvor auf Eis aufgetaut, dazupipettieren. Weitere 40 min bei 37°C inkubieren. 200 μl eiskaltes SM und einen Tropfen CHCl_3 zugeben. Bei 4°C aufbewahren oder sofort titrieren und vermehren (3.1.3.). 1 μg *in vitro* verpackte Lambda 1059 DNA ergibt bis zu 3×10^7 pfu. Durch Spalten und Religieren der DNA sinkt die Effizienz um den Faktor 10-100.

3.6.

Charakterisierung von Nukleinsäure Molekülen durch Hybridisierung

3.6.1.

Markierung von DNA durch Einbau radioaktiver Nukleotide an Einzelstrang Brüchen (Nick Translation)

Bei 0°C ansetzen:

- 1µg DNA, gelöst in TE
- 2µl 10x NT Puffer
(0.5M Tris pH 7.5, 0.1 M MgSO₄, 10mM DTT, 0.5mg/ml BSA)
- 2µl je 0.2mM dNTP
(die drei unmarkierten Nukleotide)
- 1µl 10ng/ml DNase I
(frisch verdünnt von einer 1mg/ml Stocklösung)
- 1µl DNA Polymerase I (5-10 Enzymeinheiten)
- 50µCi [α -³²P]dNTP (das markierte Nukleotid), entweder als wässrige Lösung (5µl) zugeben oder lyophilisieren und im restlichen Ansatz aufnehmen.
- H₂O ad 20µl.

(Ansätze mit mehreren markierten Nukleotiden entsprechend ansetzen.)

2 h bei 16°C inkubieren. Die Reaktion durch Zugabe von 1µl 0.5M EDTA pH 8.0 stoppen. 80µl TSE 10/100/1 pH 8.0 zugeben. Die nicht eingebauten Nukleotide durch Zentrifugation über eine 1ml Säule (spun column procedure; Maniatis et al., 1982) mit Sephadex G50 fine abtrennen (Ausschlußchromatographie).

Spezifische Radioaktivität der DNA: 5x10⁷ bis 2x10⁸ cpm/µg.

3.6.2.

Übertragung von DNA oder RNA auf Membranfilter (Southern bzw. Northern Blotting)

Das zurechtgeschnittene Agarose Gel mit der (gespaltenen und) aufgetrennten dsDNA 1 h in 1.5M NaCl, 0.5N NaOH baden. In 1M Tris pH 7.2, 1.5M NaCl neutralisieren.

Gele mit ssDNA oder RNA brauchen nicht denaturiert und wieder neutralisiert zu werden.

Ein Nitrocellulose- oder Nylon Membranfilter in der Größe des Gels zuschneiden, in H₂O und danach in 20x SSC eintauchen. Auf eine Wanne mit 20x SSC eine Plastikplatte mit Schlitzten legen. Die Enden von U-förmig gefalteten Filterpapier Streifen (z.B. aus Whatman 3MM) durch diese Schlitzte in den Puffer eintauchen. Darüber ein Filterpapier in der Größe des Gels legen. Alles mit 20x SSC durchtränken. Darauf das Gel, auf dieses das Membranfilter und darüber 2 mit 20x SSC durchtränkte Filterpapier Stücke in Gelgröße legen, ohne daß Luftblasen eingeschlossen werden. Den Aufbau mit einem Stapel Papierhandtücher bedecken. Durch Plastikfolie sicherstellen, daß der Puffer nur durch das Gel hindurch hochgesaugt werden kann. Das Ganze mit einer Glasplatte und passenden Gewichten beschweren und 6 bis 36 h (je nach Größe der zu transferierenden Moleküle) stehen lassen.

Beim Abbau der Anordnung, die Auftragekammern des Gels mit wasserfester Tinte auf das Membranfilter durchzeichnen und zugleich dessen Oberseite festlegen. Das Membranfilter 5 min in 6x SSC eintauchen. Danach an der Luft trocknen lassen.

Fixieren der Nukleinsäure Moleküle: Nitrocellulose Membranfilter 2 h bei 80°C backen; Nylonmembranfilter 2 min mit UV Licht (254nm) bestrahlen.

3.6.3.

Regenerierung bereits hybridisierter Membranfilter

Von gebrauchten Membranfiltern können die durch Hybridisierung gebundenen Nukleinsäure Moleküle durch 40 min waschen bei 90°C in PT-Waschpuffer (Thomas, 1980) entfernt werden. 12x PT-Waschpuffer: 50mM Tris pH 8.0, 2mM EDTA, 0.5% Natriumpyrophosphat, 0.02% BSA, 0.02% Polyvinylpyrrolidon, 0.02% Ficoll.

3.6.4.

Hybridisierung von filtergebundener DNA oder RNA (Southern- bzw. Northern Blot Hybridisierung)

Ein (oder mehrere; Gesamtfläche bis zu 0.1m²) Membranfilter (3.6.2.) 2 min in 6x SSC tauchen und naß in eine starke Plastikfolie (z.B. Leitz Prospekthülle Typ 4704) einschweißen. 10-30ml Vorhybridisierungslösung (50% Formamid, 5x Denhardts, 5x SSPE, 0.1% SDS und 0.4mg/ml RNA aus Hefe) einbringen. Über Nacht oder mindestens 6 h bei 42°C leicht schütteln. Die radioaktiv markierte Probe 5 min in kochendem Wasser erhitzen und zugeben. (Alternativ, die Vorhybridisierungslösung entfernen und eine bereits verwendete Hybridisierungslösung einbringen, die 5 min in kochendem Wasser erhitzt und wieder auf ca. 40°C abgekühlt wurde.) Über Nacht bei 42°C leicht schütteln. Die Hybridisierungslösung abpipettieren. Zur Kontrolle, die Radioaktivität der durch Säure fällbaren DNA der Hybridisierungslösung bestimmen.

Filter bei RT, 4mal 10 min in 2x SSC, 0.1% SDS und bei 58°C, 2mal 1 h in 1x SSC, 0.1% SDS waschen. Jeweils ca. 250ml Waschflüssigkeit nehmen. Sollte das Filter danach, an Stellen, wo sich keine eingebackenen Nukleinsäure Moleküle befinden, immer noch stark radioaktiv sein, empfiehlt sich ein weiteres Waschen bei 58°C in 0.2x SSC, 0.1% SDS.

Filter an der Luft vollständig trocknen lassen, mit Klebestreifen auf Karton gut fixieren. Einen Röntgenfilm auflegen.

Kommt es darauf an, einem radioaktiven Signal auf dem Film exakt einen Plaque auf einer Petrischale zuzuordnen, klebt man auch den Film am Karton fest und schneidet beide gemeinsam ein, sodaß sich die genaue Lage des Films während der Exposition nach der Entwicklung feststellen läßt. Die Markierungen (Nadelstiche) und die Beschriftung des Filters überträgt man auf den entwickelten Film und kann diesen dadurch mit der betreffenden Petrischale in Übereinstimmung bringen.

Expositionsdauer je nach Radioaktivität des Filters. Meistens 24 h bei -80°C .

3.6.5.

Isolierung spezifischer Klone aus einer Genbank in Lambda Phagen

450000 rekombinante Lambda Phagen einer Genbank auf 10 15cm-Petrischalen züchten (3.1.3.). Von jeder der Plaque übersäten Agaroberflächen einen oder mehrere Abdrücke nehmen.

Dazu 14cm Nitrocellulose Membranrundfilter mit einem Kugelschreiber beschriften und, ohne Luftblasen einzuschließen, auf die Agaroberfläche legen. An 3 oder mehr Stellen asymmetrisch durch das Membranfilter hindurch in der Agar stechen und diese Punkte auf der Unterseite der Petrischale mit farbiger, wasserunlöslicher Tinte markieren. Das Membranfilter nach 60 s mit einer geeigneten Pinzette vorsichtig abziehen, mit der DNA behafteten Seite nach oben auf 1.5M NaCl, 0.5N NaOH legen und nach ca. 10 s darin eintauchen. Das Membranfilter 1 min später in 0.5M Tris pH 8.0, 1.5M NaCl überführen, nach weiteren 5 min kurz in 2x SSPE schwenken und schließlich auf Filterpapier (z.B. Whatman 3MM) legen. Möchte man einen weiteren Abdruck von der selben Agaroberfläche nehmen, verfährt man in gleicher

Weise, nur daß man das Membranfilter ca. 30 s länger in Kontakt mit den Plaques läßt, an anderen Stellen einsticht und mit einer anderen Farbe markiert.

Die an der Luft getrockneten Membranfilter 2 h bei 80°C backen.

Einen Satz Membranfilter mit der ^{32}P -markierten Probe (3.6.1) hybridisieren (3.6.4.). (Die Hybridisierung eines doppelten Satzes von Membranfiltern gegen die gleiche Probe, bietet den Vorteil, daß man spezifische und unspezifische Signale - "background" - fast immer unterscheiden kann. Erstere liegen auf beiden Filtern an der gleichen Stelle, letztere sind zufällig verteilt.)

Die Radiographie der Membranfilter und die betreffenden Petrischalen mit Hilfe der Farbmarkierungen in Übereinstimmung bringen. Die Stelle an der ein Plaque sitzt, dessen DNA mit der Probe hybridisiert hat, mit einer (sterilen) Plastiköse ausstechen. Dieses Agarstückchen in 0.5ml SM einbringen und mindestens 4 h bei 4°C stehen lassen.

Die aus dem Agarstückchen herausgespülten Phagen titrieren (3.1.3.2). Von der Agarplatte (der Titration), die ca. 200 Plaques aufweist, einen Membranfilter Abdruck nehmen und, durch erneutes Hybridisieren mit der Probe, einzelne, positive Plaques isolieren. Einen (oder mehrere) dieser positiven Plaques vermehren (3.1.3). Durch eine weitere Hybridisierung testen, ob der Phagenstock plaquerein ist.

3.6.6.

Komplementaritätstest von M13 Inserts

Je ca. 0.1µg ssDNA von 2 rekombinanten M13 Phagen mit potentiell komplementären Inserts in 0.25M NaCl 1 h bei 65°C halten. Sind die Inserts komplementär, hybridisieren sie miteinander. Die Hybridmoleküle wandern bei der Gelelektrophorese weniger weit als die ssDNAs.

3.7.

DNA Sequenzierung nach der Kettenabbruch Methode

Nach Sanger et al. (1977), unter Verwendung von [α - 35 S]Thio-dATP (Biggin et al., 1983).

3.7.1.

Vorbereitung der Probe (Template/Primer)

1.5µg rekombinante Plasmid DNA (Template) in 25µl 0.2N NaOH, 2mM EDTA aufnehmen. 5 min bei RT stehen lassen. 25µl H₂O, 33µl NH₄OAc pH 4.5 und 166µl EtOH zugeben (Chen & Seeburg, 1985). 10 min bei -80°C halten, abzentrifugieren, mit 70% EtOH waschen und kurz lyophilisieren. DNA in 8µl H₂O aufnehmen, 1µl 10x Polymerase Puffer (0.6M NaCl, 0.5M Tris pH 7.5, 70mM MgCl₂, 70mM DTT) und 10pMol Primer Oligonukleotid in 1µl TE zugeben. ODER

5µl (ca. 0.5µg) einer Präparation rekombinanter M13 ssDNA (Template), 1µl 10x Polymerase Puffer, 1pMol Primer Oligonukleotid in 1µl TE und H₂O ad 10µl in ein Mikrofuge Röhrchen pipettieren.

Die Probe (Template/Primer) 5 min in kochendem Wasser erhitzen und langsam, ca. 45 min, stehend in einem mit heißem Wasser gefüllten Reagenzglas, abkühlen lassen. Dadurch wird die Anlagerung des Primer Oligonukleotids an das Template erreicht.

Primer Oligonukleotide(*):

GTAAAACGACGGCCAGT, universeller Primer für die Sequenzierung in

(*) Die Primer Oligonukleotide waren ein Geschenk von Dr. R. Mertz, Genzentrum München. Sie sind aber auch kommerziell erhältlich.

pUC und M13mp, identisch mit einer Sequenz des kodierenden Strangs des lacZ Gens neben dem Polylinker, komplementär zum M13mp (+)-Strang.

CAGGAAACAGCTATGAC, reverser Primer für die Sequenzierung in pUC, komplementär zu einer Sequenz des kodierenden Strangs des lacZ Gens neben dem Polylinker.

3.7.2.

Sequenzierungsreaktion

Die Probe und 1µl (5 Enzymeinheiten) Klenow Fragment der DNA Polymerase I zu 2µl Desoxyadenosin 5'-(α -[35 S]thio)triphosphat (650Ci/mMol; 7.5mCi/ml) pipettieren.

Je 3µl dieses Ansatzes auf 2µl A-, C-, G- und T Mix pipettieren. Die 4 Teilreaktionen 20 min bei 30°C inkubieren. Je 6µl Farbmix (96% Formamid, 10mM EDTA, 0.3% Bromphenolblau, 0.3% Xylencyanol) zugeben. 5 min im kochenden Wasserbad erhitzen. Auf Eis abkühlen.

Ca. 2µl jedes Reaktionsgemisches auf ein Sequenzgel auftragen. Der Sequenzierungsansatz kann, aufbewahrt bei -20°C, auch noch einige Tage später der Gelelektrophorese unterworfen werden.

Nukleotid Lösungen für die Sequenzierungsreaktionen:

A Mix = 0.1mM dCTP, 0.1mM dGTP, 0.1mM dTTP, 0.125mM ddATP

C Mix = 0.01mM dCTP, 0.1mM dGTP, 0.1mM dTTP, 0.5mM ddCTP

G Mix = 0.1mM dCTP, 0.01mM dGTP, 0.1mM dTTP, 0.5mM ddGTP

T Mix = 0.1mM dCTP, 0.1mM dGTP, 0.01mM dTTP, 1mM ddTTP

3.7.3.

Gel Elektrophorese

(Unter Verwendung des BRL MakrophorTM Elektrophoresesystems.)
Die hintere, thermostatisierbare Glasplatte mit Siliconlösung silikonisieren. Die vordere Glasplatte mit "Bindesilan" (gamma-Methacryloxypropyltrimethoxysilan) behandeln.

Ein 60cm langes, 0.1mm (oben) bis 0.4mm (unten) dickes Polyacrylamid Keilgel (6% Acrylamid/N,N'-Methylenbisacrylamid = 19:1 und 8M Harnstoff in TEB Puffer, polymersiert durch Zugabe von 0.06% Ammoniumpersulfat und 0.03% TEMED) gießen. Die Keilform wird dadurch erreicht, daß man die Abstandhalter ab der zweiten Hälfte des Gels zuerst doppelt, dann dreifach und schließlich vierfach übereinander legt.

Gel in die Apparatur einspannen und auf 55°C thermostatisieren. 1 h lang 2000 V anlegen. Einen "Haifischzahn" Gelkamm einsetzen. Proben auftragen. Die Elektrophorese bei ca. 2500 V und maximal 30 mA durchführen und nach ca. 2 1/2 h, wenn das Bromphenolblau den unteren Rand des Gels erreicht hat, abbrechen. (Will man lange Reaktionsprodukte noch etwas besser auftrennen, läßt man eine zweite Elektrophorese des Sequenzierungsansatzes entsprechend länger laufen.)

Das Gel, haftend an der vorderen Glasplatte, abnehmen, 30 min in 5% HOAc baden, bei 80°C vollständig trocknen und damit einen Röntgenfilm bei -80°C ca. 24 h exponieren.

4.

ERGEBNISSE

4.1.

Genbank der TC7 Zelllinie

Aus einer Genbank kann man prinzipiell jedes Gen und jeden untranskribierten Abschnitt eines Genoms isolieren. Proben für die Suche, das Durchmustern einer Genbank, stehen heutzutage, nicht zuletzt durch die chemische DNA Synthese, in großem Umfang zur Verfügung. Die *Cercopithecus aethiops* Zelllinie TC7 (siehe 2.1.) ist permissiv für das Simian Virus 40. Eine Genbank davon anzulegen bietet deshalb auch noch die Perspektive, zelluläre Gene zu isolieren, deren Expression durch die SV40 Infektion induziert oder spezifisch blockiert wird.

1.3µg statistisch mit Sau3A geschnittene, 12-20 kb große TC7 DNA (3.3.1.; 3.4.1.) wurde mit 2.3µg gereinigten Lambda 1059 Vektorarmen (3.3.5.1.; 3.4.2.) ligiert (3.5.1.1.) und danach in 6 Ansätzen *in vitro* verpackt (3.5.3.). Dies ergab 1.6×10^6 rekombinante Phagen, entsprechend einer Ausbeute von 7×10^5 pfu/µg Vektor DNA.

66x 20000 pfu davon wurden auf Q359 vermehrt (3.1.3.1.). Jeder der 66 Ansätze wurde titriert (3.1.3.2), und es wurde eine Stammlösung (I) gemischt, die aus jedem Vermehrungsansatz gleichviele Phagen enthielt.

Einen Teil dieser Stammlösung habe ich auf 5% Dimethylsulfoxid eingestellt, in Portionen zu 5×10^5 pfu aufgeteilt, durch eintauchen in flüssigen Stickstoff gefroren und bei -80°C aufbewahrt. Der andere Teil lagert bei 4°C.

Ein zweiter Ligaseansatz mit Phagen- und TC7 DNA aus derselben Präparation, der als einzigen Unterschied 7.5mM DTT enthielt, lieferte 1.3×10^6 rekombinante Phagen. Mit ihnen wurde in glei-

cher Weise weiterverfahren. Sie bilden Stammlösung II.

Plattiert man die rekombinanten Phagen für das Durchmustern wieder auf Q359 aus, so hat man eine zweimalige Selektion auf den Phänotyp Spi⁻ (siehe 1.1.). Da zudem das zentrale Lambda 1059 BamHI Fragment schon bei der Vektorpräparation abgetrennt wurde, ergibt sich eine große Sicherheit, daß praktisch alle Phagen ein Insert aus TC7 DNA tragen. So habe ich, pro 5×10^5 pfu, maximal einen Phagen gefunden, der noch das ColEI Replikon trug. Einmal habe ich diesen Phagen isoliert und dessen DNA charakterisiert. Es stellte sich heraus, daß das 5kb HindIII Fragment, welches das ColEI Replikon enthält, noch vorhanden war. Es lag aber, wie die Restriktionsanalyse zeigte, an einer anderen Stelle im veränderten Insert. Wahrscheinlich hat in diesem Fall ein Rekombinationsereignis stattgefunden, das das ColEI Replikon erhalten, die Gene für den Spi⁺ Phänotyp dagegen eliminiert hat.

Die gute Ausbeute der Ligaseansätze und das funktionierende Selektionssystem berechtigen zu der Aussage, daß es sich bei den 1.3×10^6 potentiell verschiedenen, rekombinanten Phagen der Stammlösung I und auch bei der Stammlösung II um eine TC7 Genbank handelt. Aus 1.3×10^6 rekombinanten Phagen mit durchschnittlich 16 kb großen Inserts sollte sich, statistisch gesehen (Clarke & Carbon, 1976), mit einer Wahrscheinlichkeit von 99.9%, jeder beliebige Genomabschnitt isolieren lassen.

4.2.

Das c-myc Gen der TC7 Zelllinie

4.2.1.

Isolierung und Restriktionskartierung

Beim Durchmustern (3.6.5.) von 12x 45000 Plaques der TC7 Genbank mit dem klonierten Exon 2 (1.5kb SacI Fragment in pUC12; siehe Abb. 2) des menschlichen c-myc Gens als radioaktiver Probe, fand ich 10 stark hybridisierende, rekombinante Phagen. Eine erste Restriktionsanalyse der DNA dieser Phagen zeigte, daß 9 davon identisch sind und der zehnte ein zum größten Teil gleiches Intron trägt. Die identischen Klone wurden L59l und L59li bis L59lp, der zehnte Phage wurde L592 genannt.

Southern Blot Hybridisierungsexperimente (3.6.4.) und die Restriktionskartierung der Inserts von L59l und L592 (Abb. 1) zeigten, daß das komplette c-myc Gen der TC7 Zelllinie isoliert worden war.

Im Vergleich der Restriktionskarten von L59l und L592 ist zu sehen, daß 14 kb des 15kb Inserts von L592 in der Lage der untersuchten Schnittstellen vollständig mit dem L59l Insert übereinstimmen. Dagegen ist der Abschnitt zwischen dem rechten Phagenarm und der benachbarten SacI Stelle des Inserts unterschiedlich. Er hat, im Gegensatz zu L59l, eine PvuII- aber weder eine PstI- noch eine SacI Schnittstelle. Möglicherweise stammen die Inserts von L59l und L592 von den beiden c-myc Allelen. Die Restriktionskarte des bekannten menschlichen c-myc Genomabschnitt gleicht im fraglichen Bereich der des L59l Inserts.

Das 8.1kb HindIII/EcoRI Fragment von L59l wurde in pUC12 kloniert, pHK8.1 genannt und kartiert (Abb. 2). Das entsprechende Segment der menschlichen DNA ist gut charakterisiert und vollständig sequenziert (Gazin et al., 1984).

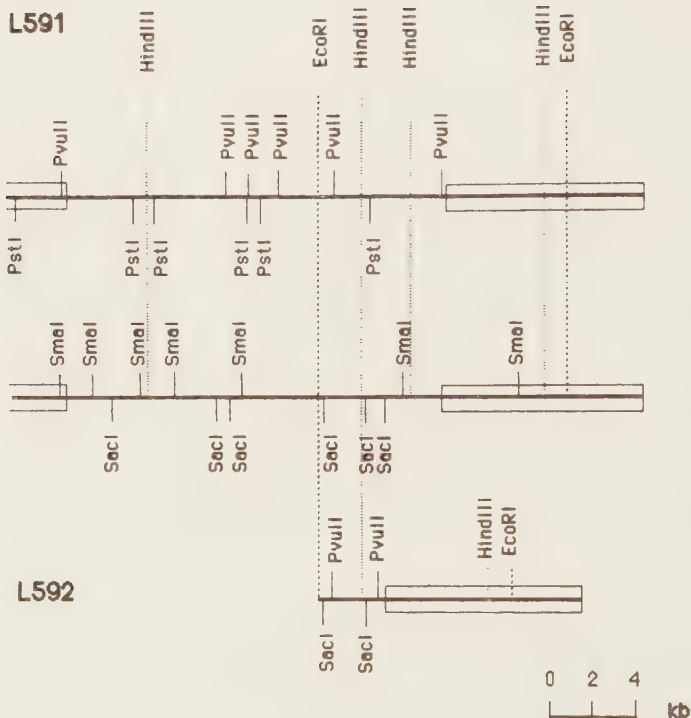


Abbildung 1: Restriktionskarte der rekombinanten Phagen mit dem c-myc Gen der TC7 Zelllinie

Im oberen Teil der Abbildung ist das 18kb Insert des rekombinanten Phagens L591 samt der flankierenden Vektorbereiche (offene Balken) dargestellt. Die Schnittstellen für die Restriktionsenzyme EcoRI, HindIII, PstI, PvuII, SacI und SmaI sind angegeben. Der untere Teil zeigt die DNA des rekombinanten Phagens L592 von der EcoRI Stelle des Inserts bis zum rechten Ende. Der Bereich zwischen der rechten SacI Stelle und dem Phagenarm unterscheidet sich von L591. Er trägt eine PvuII-, aber keine PstI- und keine SacI Stelle. Weder L591 noch L592 enthalten eine BamHI Stelle.

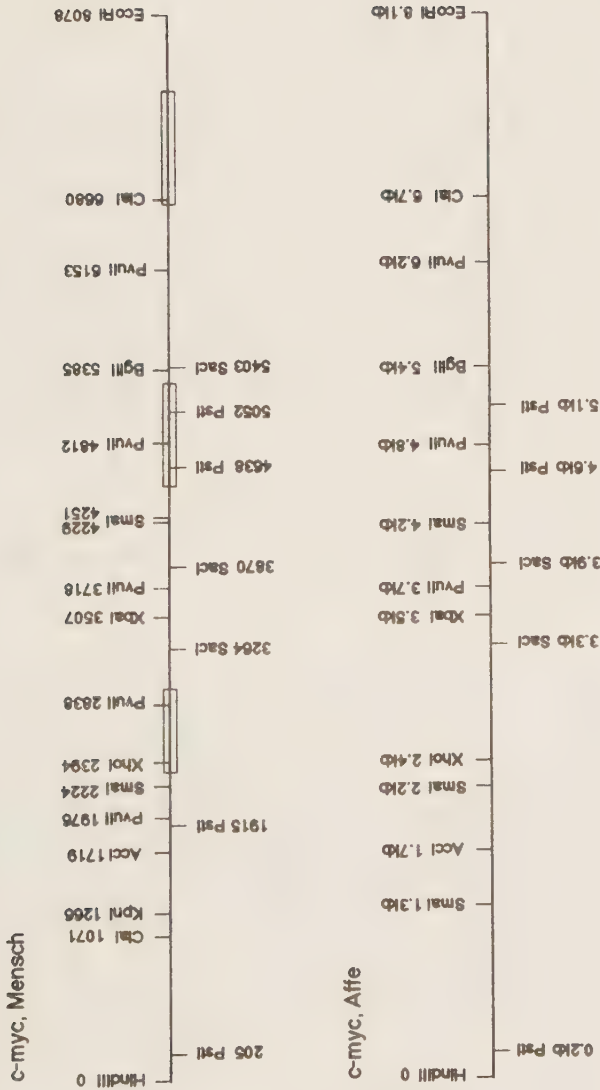


Abbildung 2: Vergleich der Restriktionskarten der c-myc Gene von Mensch und Affe. Die 8.1kb HindIII/EcoRI Fragmente der c-myc Gene von Mensch und Affe unterscheiden sich in einigen Schnittstellen: ClaI (1071), KpnI (1266), SmaI (1.3kb), PstI (1915), PvuII (1976 und 2838) und SacI (5403). Offene Balken geben die 3 Exons an.

Deshalb habe ich, in Analogie dazu, der HindIII Stelle von pHK8.1 die Position 0, der EcoRI Stelle die Position 8.1kb zugeschrieben. Auf dem 8.1kb Fragment liegen die 3 Exons und 2.3 kb des 5'-untranskribierten Bereichs. Weitere 4 kb des 5'-untranskribierten Bereichs liegen auf dem 7.7kb PvuII Fragment von L591 (siehe Abb. 1), das in pUC12 (SmaI) kloniert wurde. Die dadurch erhaltenen Klone wurden entsprechend der beiden Orientierungsmöglichkeiten ihrer Inserts p8P1 und p8P2 genannt.

Die Restriktionskarten der c-myc Gene von Mensch und Affe sind sich sehr ähnlich (Abb. 2). Größere Unterschiede gibt es nur im Exon 1 Bereich. So hat dieser Abschnitt der TC7 DNA im Vergleich zum Menschen 2 PvuII-, 1 ClaI- und 1 PstI Schnittstelle weniger, und das XhoI/SacI Fragment von 2.4-3.3 kb ist um ca. 50 bp länger.

Diese Unterschiede ließen eine Sequenzierung des Exon 1 Bereichs interessant erscheinen, um - im Vergleich mit den bereits bekannten c-myc Genen - Hinweise auf regulatorische Elemente der DNA zu bekommen.

4.2.2.

Subklonierungen

Als Basis für eine Sequenzierung und für zukünftige Experimente bezüglich der Genregulation von c-myc wurden von pHK8.1 (Abb. 2) folgende Subklone konstruiert:

pHK2.1 und m2.1: PstI/XhoI Fragment von 0.2-2.4 kb in pUC12 (SalI/PstI) bzw. in M13mpl1

pHK3.0: PstI/SacI Fragment von 0.2-3.3 kb in pUC 12

pK31 und pK32: SmaI Fragment von 1.3-2.2 kb in pUC 12 mit beiden Orientierungen

mF1 und mF2: dasselbe in M13mpl9

pHK1.2 und mA5: SmaI/SacI Fragment von 2.2-3.3 kb in pUC12 bzw. M13mpl0

pHK0.9: XhoI/SacI Fragment von 2.4 bis 3.3 kb in pUC 12 (SacI/SalI)

m10.9 und m11.9: dasselbe in M13mpl0 bzw M13mpl1

pHK1.2 wurde gezielt weiter subkloniert, indem zunächst durch Spaltung mit EcoRI und HindIII und anschließender Gel Elektrophorese das Insert vom Vektor getrennt und durch Bindung an NA45 Membranfilter (3.2.6.) isoliert wurde. Das 1.2kb Insert wurde zum einen mit AluI, zum anderen mit TaqI gespalten. Die dadurch entstandenen Fragmente wurden wiederum aufgetrennt und einzeln isoliert. Die Klonierung der AluI Fragmente erfolgte in pUC12 (SmaI) bzw. in M13mpl9, die der TaqI Fragmente in pUC12 (AccI) bzw. M13mpl9. Die rekombinanten M13 Klone wurden bezüglich der Orientierung ihrer Inserts durch Komplementaritätstests (3.6.6.) charakterisiert.

4.2.3.

Sequenzierung

Nachdem sich gezeigt hatte, daß die Sequenzierung (3.7.) nach Sanger et al. (1977) mit M13 ssDNA bessere Ergebnisse brachte als mit Plasmid DNA, habe ich die Sequenz von 2.0-3.3kb des pHK8.1 Inserts durch die Analyse ausgewählter, rekombinanter M13 Klone (Abb. 3) bestimmt.

Die ermittelte Sequenz (Abb. 4) der Exon 1 Region des TC7 c-myc Gens trägt den Namen TC7myc1. Pos. 1 entspricht dabei der Pos. 1976 (PvuII Stelle) des 8.1kb Fragments mit dem c-myc Gen des Menschen.

TC7myc1

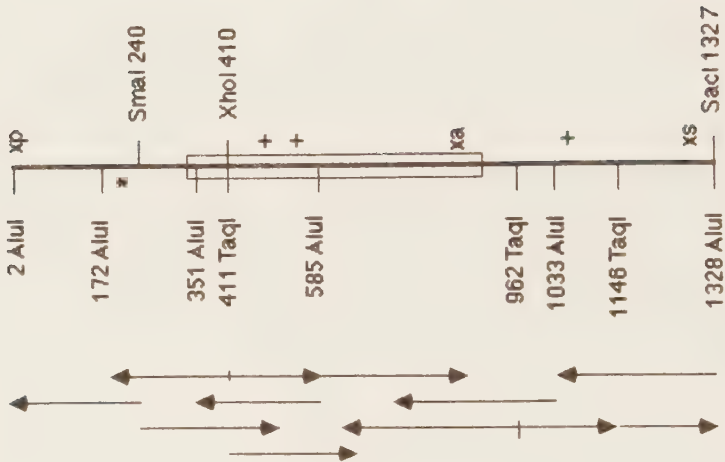


Abbildung 3: Die Sequenzierung des c-myc Exon 1 Bereichs von pHK8.1

Die Strategie für die Sequenzierung des Exon 1 Bereichs von pHK8.1 ist dargestellt. Die Pfeile bezeichnen die Länge und die Richtung der gelesenen Sequenzen. Exon 1 ist durch einen offenen Balken gekennzeichnet. Die vollständige Sequenz trägt den Namen TC7myc1. Einige Unterschiede im Vergleich zur entsprechenden menschlichen Sequenz sind vermerkt: 9bp weniger (*), 1bp mehr (+), PvuII, AluI und SacI Stelle fehlen (xp, xa und xs). Neu sind die AluI Stelle (1033) und die SacI Stelle (1327).

TC7myc1

CAGCTCTTTC	GCCTGCGATG	ATTTAAACTC	ACAGGACAAG	GATGCGGTTT	50
GTCAAAGAGT	GCTGCTACGG	AGGAGCAGCA	GAGAAAGCGA	GAGGGTTTGA	100
GCAGGAGCAA	AAGAAAATGG	TAGGCGCGCG	CAGTTAATTC	ATGCGGCTCT	150
CTTACTCTGT	TTACATCCGA	GAGCTAGAGT	GCTCGGCTGC	CGGGCTGAGT	200
CTCCTCCCCA	CCCTCCCCAC	TCTCCTCAGA	AGCGCCCTC	CCGGGTTCCT	250
AAAGCAGAGG	GCGTGCGGGA	AAAGAAAAAA	GATCCTCTCT	CGCAAAATCTC	300
CGCCACAAG	CCCTTTTATA	TGCGAGGGTC	TGCGCGGCGG	AGGACCCCGG	350
AGCTGTGCTG	CTCGCGGCCG	CCACTGCCGG	CTCCCGGCAG	TCCCTGGCTC	400
CCCTCCTGCC	TCGAGAAGGG	CAGGGCTTCT	CAGAGGCTTG	GCGGGAAAAA	450
GAACGGAGGG	AGGGATCGCG	CTGAGGTATA	AAAGCCGGTT	TTCGGGGCTT	500
TATCTAACTC	GCTGTAGTAA	TTCCAGCGAG	AGGCAGAGGG	AGCGAGCGGG	550
CGGGCCGGCG	AGGGTGGAAG	ATCCGGGCGA	GCAGAGCTGC	GCTCCGGGCG	600
TCGTGGGAAG	GGAGATCCGG	AGCGAAAAGG	GGGCTTCGCC	TCCGGCCCAG	650
CCCTCCCGCT	GACCCCCCAG	ACAGCGGTCC	GCAACCCTTG	CCGCATCCAC	700
GAAACTTTGC	CCATAGCAGG	GGGCGGACAC	TTTGCACTGG	AACTTACAAC	750
ACCCTAGCAA	GGACGCTACT	CTCCCGACGC	GGGGAGGCTA	TTCTGCCTAT	800
TTGGGGACAC	TTCCCCGCGG	CTGCCAGGAC	CCGCTCCTCT	GAAAGGCTCT	850
CCTTGCCGCT	GTTTGGACGC	TGGATTTTTT	TCCGGTAGTG	GAAAACCAGG	900
TAAGCACC GA	GGTCCATT TG	TCTTTTAATT	TATTTTTGTA	TCGCTTTAAT	950
GCTGAGATGA	GTCGAATGCC	GAAATAGGGT	GTCTTTTCTC	CCATTCTCTG	1000
GCTATTGACA	CTTTTCTCAG	AGTAGTTGTG	GTAGCTTGGG	CTGGGGTAGG	1050
GGGTTAATCC	AGAACTGGAT	CGGGATAAAG	TGACTTGTC A	AGATGGGAGA	1100
GGGAGACGGCA	GAGGGAAAAAC	GGGAATGGTT	TTTAAGACTA	GCCTTTCGAG	1150

```

ATTCTGCGT CTTGAATATA TTCACGCTGA CTCCCGGCCG GTCGGACATT 1200
CCTGCTTTAT TGGGTTAATT GCTGGCTGAG TTTTGGGGGG CTGGGGGTTG 1250
CTTTGCGGTG GGCAGAAAGC CCCTTGCATC CTGACCTCCT TGGAGTAGGG 1300
ACCGCATATT GCCTGTGTGA GCCAGAGAGC TC                      1332

```

Abbildung 4: TC7mycl, die Nukleotidsequenz des Exon 1 Bereichs des c-myc Gens der Affen Zelllinie TC7
 TC7mycl enthält 343 bp des 5'-untranslatierten Bereichs, das Exon 1 (Pos. 344-899) und einen Teil des Introns 1 des c-myc Gens und entspricht der menschlichen Sequenz von Pos. 1976 bis 3313 des 8.1kb HindIII/EcoRI Fragments mit dem kompletten c-myc Gen.

4.2.4.

Southern- und Northern Blot Analyse

Eine Southern Blot Analyse (3.6.4.) gespaltener genomischer TC7 DNA mit pHK0.9 (enthält den größten Teil des 1. Exons) als Probe, zeigte die Restriktionsfragmente: $\geq 25\text{kb}$ BamHI, $\approx 13\text{kb}$ PvuII, $\approx 12.5\text{kb}$ EcoRI, 10.1kb HindIII, 5.2kb SacI, 4.4kb PstI, 2.0kb SmaI als markierte Banden. Dieses Ergebnis steht im Einklang mit der Restriktionskarte. Die Intensität der Banden im Vergleich mit einer definierten Plasmid DNA Probe (Marker) entsprach einer c-myc Kopie pro Genom.

Die Northern Blot Analyse cytoplasmatischer RNA der TC7 Zelle mit pHK0.9 als Probe zeigte eine c-myc spezifische ca. 2.3kb Bande (Abb. 5). Dies entspricht genau der Northern Blot Analyse normaler menschlicher Zellen (ar-Rushdi et al., 1983).

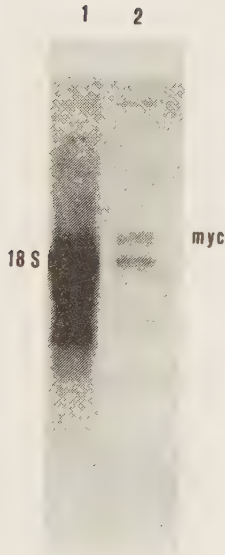


Abbildung 5: Northern Blot Analyse der TC7 Zelle mit einer c-myc Exon 1 Probe

40µg cytoplasmatische RNA (Spur 1) und 4µg 1x poly-A selektionierte RNA (Spur 2) aus TC7 Zellen (3.3.2.; 3.3.3.) wurden durch Gelelektrophorese aufgetrennt (3.2.5.2.), auf ein Nylon Membranfilter transferiert (3.6.2.) und gegen radioaktiv markiertes (3.6.1.) pHK0.9 hybridisiert (3.6.4.). Die Abbildung zeigt die Autoradiographie. Die c-myc mRNA Bande liegt bei ca. 2.3kb. Darunter (bei 1.9kb) liegt die Bande der 18S rRNA, die durch einmalige Oligo-(dT) Chromatographie nicht vollständig entfernt werden kann und deren Markierung wahrscheinlich auf der Homologie mit den G+C reichen TC7myc1 Sequenzen 536-562 und 768-787 beruht.

4.2.5.

Sequenzvergleich

Der direkte Vergleich zwischen TC7mycl und der homologen menschlichen Sequenz (Pos. 1976-3313; Gazin et al., 1984) zeigt eine Übereinstimmung von 94.2% (Abb. 6).

Aufgrund dieser großen Homologie und der Northern Blot Analyse (4.2.4.) gilt als sicher, daß die TC7mycl Sequenz 344-899 das Exon 1 des c-myc Gens des Affen darstellt.

Damit liegen die TATA Boxen auf den Pos. 316 und 477, die cap-Stellen bei 344 und 507 und die 5'-Spleißstelle des 1. Introns auf Pos. 900 von TC7mycl.

Der einzige größere Unterschied sind 9 bp weniger bei TC7mycl Pos. 204. Ansonsten gibt es zusätzliche Basenpaare bei den TC7mycl Pos. 476, 554 und 1055 sowie 66 Pos. mit unterschiedlichen Basen. Letztere sind der Grund für die Unterschiede in den Schnittstellen. So sind z.B. AluI (Pos. 1033) und SacI (Pos. 1327) spezifisch für TC7mycl oder PvuII (Pos. 1976), PvuII (Pos. 2838), AluI (Pos. 2839) und SacI (Pos. 3264) spezifisch für die menschliche c-myc Sequenz. (Siehe auch Abb. 3.)

TC7mycl Exon 1 hat kein ATG Startkodon und keinen offenen Leserahmen. Es ist deshalb anzunehmen, daß der offene Leserahmen im Bereich des 1. Exons des menschlichen c-myc Gens (ATG 2304 - TAG 2868; Gazin et al., 1984) keine funktionelle Bedeutung hat, zumal auch dort kein ATG Kodon im 1. Exon (cap-Stelle 1: 2328; cap-Stelle 2: 2490) liegt.

Ein direkter Sequenzvergleich der homologen Bereiche von Affe (TC7mycl) und Maus (Bernard et al., 1983) ist ebenfalls möglich (Abb. 7). Sie stimmen zu 69.0% überein. Auffallend ist, daß die einander entsprechenden Pyrimidin reichen Sequenzen TC7mycl 200-227 und Maus c-myc 248-275 gleich lang sind, während das menschliche Gen an dieser Stelle 9 bp mehr aufweist.

TC7myc1

HUMAN (LAWN) C-MYC PROTO-ONCOGENE

1976-3313

1 CAGCTCTTTTCGCCTGCGATGATTTAACTCACAGGACAAGGATGCGGTTTGTCAAAGAGT
 1976 CAGCTGTTCCGCCTGCGATGATTTATACTCACAGGACAAGGATGCGGTTTGTCAAACAGT

61 GCTGCTACGGAGGAGCAGCAGAGAGAAAGCGAGAGGGTTTGAGCAGGAGCAAAAGAAAATGG
 2036 ACTGCTACGGAGGAGCAGCAGAGAGAAAGGGAGAGGGTTTGAGAGGGAGCAAAAGAAAATGG

121 TAGGCGCGCGCAGTTAATTCATGCGGCTCTCTTACTCTGTTTACATCCGAGAGCTAGAGT
 2096 TAGGCGCGCGTAGTTAATTCATGCGGCTCTCTTACTCTGTTTACATCCTAGAGCTAGAGT
 U1/2 Homologie

181 GCTCGGCTGCCGGGCTGAGTCTCC TCCCCACCTCCCCACTCTCCTCAGAA
 2156 GCTCGGCTGCCGGGCTGAGTCTCTCTCCCCACCTTCCCCACCTCCCCACCTCCCCATAA
 polyQ

232 GCGCCCCCTCCCGGGTTCCCAAAGCAGAGGGCGTGGGGGAAAAAGAAAAAGATCCTCTCTC
 2216 GCGCCCCCTCCCGGGTTCCCAAAGCAGAGGGCGTGGGGGAAAAAGAAAAAGATCCTCTCTC
 polyP

292 GCAAATCTCCGCCACAAGCCCTTTATAATGCGAGGGTCTGCGCGGCGGAGGACCCCGA
 2276 GCTAATCTCCGCCACCGGCCCTTTATAATGCGAGGGTCTGGACGGCTGAGGACCCCGA
 Spl Homologie ▶ cap 1

352 GCTGTGCTGCTCGGGGCCGCCACTGCCGGCTCCCGGCAGTCCCTGGCTCCCTCCTGCCT
 2336 GCTGTGCTGCTCGGGGCCGCCACCGCCGGGCCCGGCCGTCCCTGGCTCCCTCCTGCCT

412 CGAGAAGGGCAGGGCTTCTCAGAGGCTTGGCGGGAAAAAGAACGGAGGGAGGGATCGCGC
 2396 CGAGAAGGGCAGGGCTTCTCAGAGGCTTGGCGGGAAAAAGAACGGAGGGAGGGATCGCGC
 hsp83 Homologie polyP

472 TGAGGTATAAAAGCCGGTTTTTCGGGGCTTTATCTAACTCGCTGTAGTAATTCCAGCGAGA
 2456 TGA GTATAAAAGCCGGTTTTTCGGGGCTTTATCTAACTCGCTGTAGTAATTCCAGCGAGA
 ▶ cap 2


```

1192 TCGGACATTCTCTGCTTTATTGGGGTTAATTGCTGGCTGAGTTTTGGGGGGCTGGGGGTTGC
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3173 TCGGACATTCTCTGCTTTATTGTGTTAATTGCTCTCTGGGTTTGGGGGGCTGGGGGTTGC

1252 TTTGCGGTGGGCAGAAAGCCCCTTGCATCCTGACCTCCTTGGAGTAGGGACCGCATATTG
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3233 TTTGCGGTGGGCAGAAAGCCCCTTGCATCCTGAGCTCCTTGGAGTAGGGACCGCATATCG

1312 CCTGTGTGAGCCAGAGAGCTC
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3293 CCTGTGTGAGCCAGATCGCTC

```

Matches = 1263 Mismatches = 66 Unmatched = 12
 Length = 1341 Matches/length = 94.2 percent

Abbildung 6: Sequenzvergleich zwischen TC7mycl und der dazu homologen Region des c-myc Gens des Menschen.

Vergleich der TC7mycl Sequenz mit der Sequenz 1976-3313 des menschlichen c-myc Gens (Gazin et al., 1984) unter Verwendung des Microgenie™ "ALIGN" Programms (Queen & Korn, 1984). TATA Boxen, cap-Stellen, die 5'-Spleißstelle des 1. Introns und einige potentielle regulatorische Elemente (siehe 4.2.6.) sind hervorgehoben

TC7mycl

MOUSE C-MYC GENE 5' FLANK, EXONS 1 AND 2

55-1386

```

1      CAGCTCTTTGCGCTGCGATGATTTAAACTCACAGGACAAGGATGCGGTTTGTCAAAGAGT
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55     CAGCTGTTCCACCTGCGGT ACTGATATACGCAGGGCAAGAACACAGTTTCAGC  CGAGC

61     GCT GCTACGGAGGAGCAGCAGAGAAAGCGAGAGGGTTTGAGCAGGAGCAAAAGAAAATG
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112    GCTGGCGCCCGAACAACCGTACAGAAAGGGAAGGACTAGCGCGCGAG AAGAGAAAATG

120    GTAGGCGCGCGCAGTTAATTCATGCGGCTCTCTTACTCTGTTTACATCCGAGAGCTAGAG
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171    GTCGG GCGCGCAGTTAATTCATGCTGCGCTATTA CTGTTTACACCCCGAGCCGGAG

180    TGCTCGGCTGCCGGGCTGAGTCTCCTCCCCACCCTCCCCACTCTCCTCAGAAGCGCCCTT
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228    TACTGGGCTGCGGGGCTGAGGCTCCTCCTCCTCTTTCCCCGGGCTCCCCACTAGC CCCCT

```

240 CCCGGGTTCCCAAAGCAGAGGGCGTG GG GGAAA AGA
 287 CCCGAGTTCCCAAAGCAGAGGGCGGGGAAACGAGAGGAAGGAAAAAATAGAGAGAGGTG
 276 AAAAAAGATCCTCTCTCGCAAATCTCCGCCACACAAGCCCTTTATAATGCGA
 347 GGGAAAGGAGAAAAGAGAGGTTCTCTGGCTAATCCCGGCCACCCGCCCTTTATATTCCGG
 326 GGGTCTGCGCGGCGGAGGACCCCCGAGCTGTGCTGCTCGCGGCCGCCACTGCCGGCTCCC
 407 GGGTCTGCGCGGCCGAGGACCCCTG GCTGCGCTGCTCTCAGCTGCCGGGTCCGACTCGC
 386 GGCAGTCCCTGGCTCCCCTCCTGCCTCGAGAAGGGCAGGGCTTCTCAGAGGCTTGCGGGG
 466 CTCACTC AGCTCCCCTCCTGCCTCCTGAAGGCGAGCG TTCGCCGACGCTTGCGGGG
 446 AAAAAAGAACGGA GGGAGGGATC GCGCTGAGGTATAAAAGCCGGTTTTCGG GGCTT
 522 AAAAAAGAGGGAGGGGAGGGATCCTGAGTCGCAGTATAAAAGAAGCTTTTCGGGCGTTTT
 501 TATCTAACTCGCTGTAGTAATTCCAGCGAGAGGCAGAGGGAGCGAGCGGGCGGGCCGGCG
 582 TTTCTGACTCGCTGTAGTAATTCCAGCGAGAGACAGAGGGAGTGAGCGGAC
 561 AGGGTGGAAGATCCGGGCGAGCAGAGCTGCGCTCC GGGCGTGTGGGAAGGGAGATCCG
 633 GGTGGAAGAGCCGTGTGTGCAGAGCCGCGCTCCGGGGCGACCTAAGAAGGCAGCTCTG
 620 GAGCGAAAAGGGGGCTTCGCCTCCGGCCAGCCCTCCCGCTGACCC CCCAG
 692 GAGTGAGA GGGGCTTTGCCTCCGAGCCTGCCCGCCCACTCTCCCCAACCCTGCGACTG
 671 ACAGC G GTCCGCAACCCTTGCCGCATCCACGAACTTTGCCCATAGCAGGGGG
 750 ACCCAACATCAGCGGCCGCAACCCTCGCCGCCGCTGGGAACTTTGCCCATTCAGCGGG
 724 CGGACACTTTGCACTGGAACCTTACAACACCCTAGCAAGGACGCTACTCTCCCGAC GCGG
 810 CAGACACTTCTCACTGGAACCTTACAATCTGCGAGCCAGGACAGGACTCCCCAGGCTCCGG
 783 GGAGGCTA TTCTGCCTATTTGGGGACA CTTCCCCGCCGCTGCCAG GACCCGCTCCT
 870 GGAGGGAATTTTGTCTATTTGGGGACAGTGTCTCTGCCTCTGCCCGCGATCAGCTCTC
 839 CTGAAAGGCTCTCCTTGCCGCTGTTTGGACGCTGGATTTTTTCCGGTAGTGAAAAACCA
 930 CTGAAAAGAGCTCCTCG AGCTGTTTGAAGGCTGGATTTCTTTGGGCGTTGGAAACCC
 899 GGTAAGCACCGAGGTCCATTTGTCTTTTAATTTATTTTGTATCGCTTTAATGCTGAGAT
 989 GGTAAGCAC AGATCTGGTGCTTTTCCCTGTGTCTT TCT GC GTCTTGAATT

```

959  GAGTCGAATGCCGAAATAGGGTGTCTTTTCTCCCATTCCTGCGCTATTGACACTTTTCTC
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1041 GTAGCG   GCCG GTTAGGACAGTCTTTCTTCCATTCTGTGCTTTTGACACTTTTCTC
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1019  AGAGTAGTTGTGGTAGCTTGGGCTGGGGTAGGGGGTTAATCCAGAACTGGATCGGGATA
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1097  AAGAGTAGTTGGGGTA   GGCTGGGGT   AGATCTGAGTCGGGGTA
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1078  AAGTGACTTGTCAGATGGGAGAGGAGACGGCAGAGGGAAAACGGGAATGGTTTTTAAGA
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1139  GAGCGACTTGTCAGATGACAGAGGAAAGGGGA AGGGAAAACCGGGATGCATTTTGAAG
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1138  CTAGCCTTTTCGAGATTTCTGCGTCTTGAATATATTACGCTGACTCCCGCGCGGTTCGGAC
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1198  CGGGGTTCCCGAGGTTACT   ATGGGCTGACG TGA   CCCGCGCGGTGGAC
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1198  ATTCCTGCTTTATTGGGTTAATTGCTGGCTGAGTTTTGGGGG   GCTGGG GGTGCTT
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1246  ATTCTTGCTTTGCTACAGGAATTGATATGTGCCTTTGAGGGGCAAACCGGGAGGTCGCTT
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1254  TGCGGTGGGC   AGAAAGCCCCTTGCACTCCTGACCTCCTTGAGTAGGGACCGCATATTG
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1306  CGTGGTGGCCAAAGAAAGCCCTTGGAATCCTGAGGTCTTTGGAGAAGGGATTACCTTTG
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1312  CCTGTGTGAGCCAGAGAGCTC
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
1366  CGTTTGGGAGCGAGAAGGCTC

```

Matches = 966 Mismatches = 297 Unmatched = 138
Length = 1401 Matches/length = 69.0 percent

Abbildung 7: Sequenzvergleich zwischen TC7mycl und der dazu homologen Region des c-myc Gens der Maus.

Vergleich der TC7mycl Sequenz mit der Sequenz 55-1386 des c-myc Gens der Maus (Bernard et al., 1983) unter Verwendung des Microgenie™ ALIGN Programms (Queen & Korn, 1984).

Dagegen ist die Polypurin Sequenz im 5'-untranslatierten Bereich des c-myc Gens bei der Maus (Pos. 298-363) erheblich länger als bei den Primaten.

Die TC7mycl entsprechende Region des c-myc Gens des Huhns (Shih et al., 1984; siehe Anhang 7.1.1.) ist schon so verschieden, daß die größte Homologie zu TC7mycl aus einer 23bp Sequenz mit

82.6% Übereinstimmung besteht.

Schlüsselt man die Homologie zwischen Affe und Mensch oder Affe und Maus nach der Funktion der einzelnen Genabschnitte auf, sieht man, daß Exon 1 etwas besser konserviert ist als der 5'-untranskribierte Bereich und der Teil des Introns (Tab. 1), mit der größten Übereinstimmung im Bereich zwischen den beiden cap-Stellen (Affe/Mensch 97.0% - Affe/Maus 73.8%).

Ein ähnliches Bild vom Ausmaß der Homologie dieser Genabschnitte von Affe, Mensch, Maus und Huhn ergibt der Vergleich der Basenzusammensetzung (Tab. 2): kleine Unterschiede zwischen Mensch und Affe, die über die Maus hin zum Huhn zunehmen.

Das Exon 1 des menschlichen N-myc Gens (Kohl et al., 1986; siehe Angang: 7.1.2.) ähnelt in der Basenzusammensetzung dem Exon 1 der c-myc Gene von Affe, Mensch und Maus.

Homologie				
TC7mycl	1-343	91.8%	1976-2327	c-myc, Mensch
	344-899	95.7%	2328-2881	
	900-1332	94.2%	2882-3313	
TC7mycl	1-343	68.3%	55-424	c-myc, Maus
	344-899	72.4%	425-989	
	900-1332	65.0%	990-1380	

Tabelle 1: Ausmaß der Homologie funktioneller c-myc Genabschnitte zwischen Affe, Mensch und Maus

Das Ausmaß der Homologie zwischen TC7mycl und den dazu homologen Sequenzen von Mensch und Maus ist nach der Funktion der Genabschnitte aufgeschlüsselt. Bei TC7mycl: 1-343 = 5'-untranskribierter Bereich; 344-899 = Exon 1; 900-1332 = Teil des 1. Introns.

	Sequenz	%G+C	%P	%CG	
TC7mycl	1-343	56.3	52.8	6.1	5'-untranskr.
	344-899	63.8	51.1	8.3	Exon 1
	900-1332	49.4	50.5	3.7	Intron 1 (Teil)
c-myc, Mensch	1976-2327	56.3	50.6	5.1	5'-untranskr.
	2328-2881	64.4	51.1	8.5	Exon 1
	2882-3313	48.4	49.5	3.2	Intron 1 (Teil)
c-myc, Maus	55-424	59.7	55.5	7.3	5'-untranskr.
	425-989	60.5	49.2	6.6	Exon 1
	990-1386	51.9	53.2	4.0	Intron 1 (Teil)
c-myc, Huhn	850-1203	71.8	41.8	13.3	5'-untranskr.
	1204-1540	62.3	40.1	12.8	Exon 1
	1541-1950	62.9	51.4	8.8	Intron 1 (Teil)
N-myc, Mensch	1-589	62.8	49.8	7.8	Exon 1

Tabelle 2: Vergleich der Basenzusammensetzung untranslatierter Bereiche der c-myc Gene von Affe, Mensch, Maus und Huhn und des N-myc Gens des Menschen.

Die Parameter G+C-Gehalt, Purinanteil (%P) des U-Strangs und Häufigkeit des Dinukleotids CG von TC7mycl und den entsprechenden Bereichen des Menschen (Gazin et al., 1984), der Maus (Bernard et al., 1983) und des Huhns (Shih et al., 1984) sowie vom Exon 1 des menschlichen N-myc Gens (Kohl et al., 1986) sind miteinander verglichen.

Die Konservierung der kodierenden Nukleotidsequenz des c-myc Gens ist erheblich größer. Zwischen Mensch und Maus bestehen 89% (Bernard et al., 1983), zwischen Mensch und Huhn oder MC29 Virus (Watson et al., 1983b) ca. 70% Homologie.

Trotz der zum Teil erheblichen Sequenzunterschiede der verglichenen nicht translatierten c-myc Regionen ist die Genstruktur als solche konserviert. Sie findet sich auch beim verwandten N-myc Gen. Gemeinsamkeiten sind: Größe und Zahl der Exons/ Introns, mindestens 2 cap-Stellen, ein nicht translatierbares Exon 1, entsprechend einer ungewöhnlich langen 5'-Führungssequenz (leader, 5'-untranslatierter Bereich) und eine Translationsstartstelle im vordersten Teil des 2. Exons (Abb. 8).

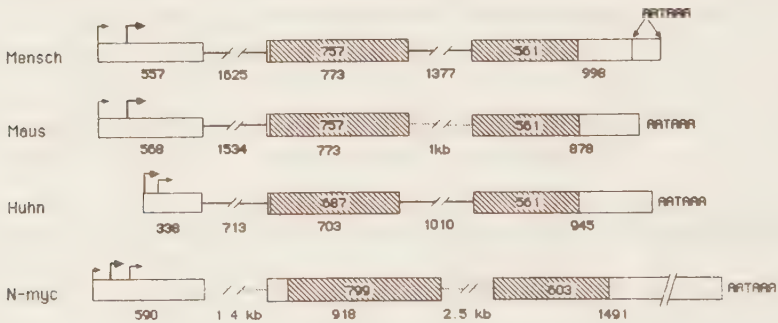


Abbildung 8: c-myc/N-myc Genstruktur

Die Exons (nicht kodierende Bereiche: offene Balken; kodierende Bereiche: gestrichelte Balken) und Introns der c-myc Gene des Menschen, der Maus und des Huhns und des N-myc Gens des Menschen sind schematisch dargestellt und mit Größenangaben versehen. (Dicke) Pfeile bezeichnen die (hauptsächlich verwendeten) cap-Stellen (Battey et al., 1983; Bernard et al., 1983; Linial & Groudine, 1985; Kohl et al., 1986). Punktuelle Linien bezeichnen nicht sequenzierte Abschnitte.

4.2.6.

Potentielle regulatorische Elemente des c-myc Gens

Die c-myc Gene von Mensch und Maus besitzen zwei Promotoren (Abb. 8), genannt P1 und P2, von denen P2 in der normalen Zelle die weit größere Aktivität hat (Battey et al., 1983; Taub et al., 1984a).

Die Expression der bakteriellen Chloramphenicol Acetyltransferase (CAT) nach Transfektion rekombinanter Plasmide mit der Promotorregion des menschlichen c-myc Gens (M. Lipp, R. Schilling, S. Wiest, G. Laux und G. W. Bornkamm, zur Publikation eingereicht) hat gezeigt, daß insbesondere P2 (cap-Stelle 2, Pos. 2490, Abb. 6) ein starker Promotor ist. Daneben konnten 2 die Transkription aktivierende DNA Abschnitte des menschlichen c-myc Gens charakterisiert werden: Das PvuII/SmaI Fragment (Pos. 1976-2224) und das XhoI/NaeI Fragment (Pos. 2394-2536).

Die Untersuchung der Promotorregion der myc Gene auf Repetitionen, auf besonders gut konservierte Bereiche und auf Homologie mit Promotor- oder Enhancerelementen (Anhang 7.1.3. und 7.2.) ergab einige interessante Aspekte.

4.2.6.1.

Das Sequenzmotiv GTTACAT

Für die effiziente Transkription der U1 oder U2 snRNA Gene (sn = small nuclear) ist eine DNA Sequenz mitverantwortlich, die ca. 220 bp vor der Transkriptionsstartstelle liegt (Murphy et al., 1982; Westin et al., 1984). Dieses Enhancerelement (Ares et al., 1985; Mattaj et al., 1985) enthält ein Sequenzmotiv (SphI Motiv; Anhang 7.2.3.) mit der Konsensussequenz ATGQAAAQ (Krol et al., 1985), welches auch im SV40 Enhancer, im IgH Enhancer und in den Promotoren der Immunglobulin Gene

vorkommt. Im Falle der U1/U2 snRNA Gene (Ares et al., 1985) und eines V_H-Promotors (Mason et al., 1985) ist die Wichtigkeit dieses Sequenzmotivs für die Transkription bereits nachgewiesen worden. Und, Singh et. al (1985) haben gezeigt, daß ein Protein aus dem Kern menschlicher B-Zellen oder von HeLa Zellen spezifisch an diese Promotor- und Enhancerelemente der Immunglobulin Gene bindet. Das Zentrum des geschützten Bereichs im DNase I Test (footprint) ist das konservierte Oktanukleotid ATTTGCAT.

Im c-myc Gen findet sich in vergleichbarer Position das Sequenzmotiv GTTTACAT bei Mensch und Affe, GTTTACAC bei der Maus und GTTTACTT beim Huhn (Abb. 9). Bei Mensch und Affe

2130	ctctGTTTACATcct	2144	c-myc, Mensch, -198 cap1, -360 cap2
155	ctctGTTTACATcgg	169	c-myc, Affe, -189 cap1, -352 cap2
203	tactGTTTACACccc	217	c-myc, Maus, -222 cap1, -385 cap2
845	ctatGTTTACTTccg	859	c-myc, Huhn, -359 cap
-251	gcgcGTTTACATacg	-265	Ula Gen, <i>X.laavis</i>
-210	tcgaATTTGCATgcc	-224	U2 Gen, Mensch
-220	tctgATTTACATgtc	-234	U2 Gen, Ratte
-104	gctgATTTGCATgta	-90	V _H -Promotor (K2), Maus
-74	gctgATTTGCATtgg	-88	V _H -Promotor (141), Maus
129	catgCTTTGCATact	143	SV40 Enhancer
162	ggtaATTTGCATttc	176	Ig _H -Enhancer, Maus

Abbildung 9: Das Sequenzmotiv PTTTPCAT im 5-untanskribierten Bereich des c-myc Gens.

Negative Zahlen geben den Abstand der Sequenzen von der (den) cap-Stelle(n) an. Die Ula Sequenz ist aus Krol et al. 1985, die U2 Sequenzen sind aus Ares et al. (1985), die V_H- und die V_H-Promotor Sequenzen sind aus Falkner & Zachau (1984) und die Ig_H-Enhancer Sequenz ist aus Banerji et al. (1983).

beträgt die Homologie mit dem Oktanukleotid im *X. laevis* Ula Enhancer Element 100%. Das Sequenzmotiv des Huhns ist Teil der oben (4.2.5.) erwähnten 19/23 bp Homologie.

4.2.6.2.

Die repetitive Polypyrimidin Sequenz

Die Sequenz 2178-2213 des menschlichen c-myc Gens enthält repetitiv - 2x perfekt und 2x mit einem Austausch - das Nonanukleotid CCTCCCCAC. Dies ist zugleich der Bereich mit dem größten Unterschied zur TC7mycl Sequenz, denn jene hat an der entsprechenden Stelle 9 bp weniger und 2x ein T statt einem C. Doch auch die Sequenz der Maus hat in diesem Bereich 9 bp weniger. Die TC7mycl Sequenz 195-235 enthält 2x perfekt das Nonanukleotid und, wegen der beiden T, zudem eine Heptanukleotid Repetition. Im c-myc Gen des Huhns findet sich, trotz des großen generellen Unterschieds der 5'-untranslatierten Region, eine ähnliche Sequenz in vergleichbarer Distanz zur (hauptsächlich verwendeten) cap-Stelle. Und auch 251 bp 5' zur cap-Stelle des Gens für das cytoplasmatische β -Aktin des Huhns (Kost et al. 1983) liegt eine gleichartige Sequenz mit dem Nonanukleotid. Alle angeführten Sequenzen (Abb. 10) bestehen im wesentlichen aus Pyrimidinnukleotiden. Homopurin/Homopyrimidin Sequenzen sind charakteristisch für Nuklease hypersensitive Bereiche, wie sie oft in 5'-flankierenden Regionen aktiver (transkribierter), eukaryotischer Gene gefunden wurden (Evans et al., 1984; kurzer Übersichtsartikel: Elgin, 1984).

Die c-myc Homopyrimidin Folge zeigt große Homologie zu der S1 Nuklease hypersensitiven Stelle (Evans et al., 1984) des SV40 Enhancers, einem Bereich der 72bp Repetition, der das für die Enhancerwirkung wichtige Sequenzmotiv TCCCCA (Zenke et al., 1986) 2x enthält.

Von Siebenlist et al. (1984) ist die Lage der DNase I hypersen-

2170 ctgagtCTCTCCCGACCTTCCCGACCGTCCCGACCGTCCCGATAagcg 2219 c-myc, Mensch, -271 cap2
 195 ctgagtCTCTCTCCCGAC CCGTCCCGACTCTCTCTAGAagcg 235 c-myc, Affe, -272 cap2
 243 ctgaggCTCTCTCTCTCT CTTTCCCG GCTCCCGACTagccc 283 c-myc, Maus, -305 cap2
 901 catcgCTC TCCCG GCGGCT (CTCC)s cgccgccc 950 c-myc, Huhn, -255 cap
 -257 ctgctt CACTCTCCCGA TCTCCCCCGCTCCCGACccccaa -217 β -Aktin Gen, Huhn
 172 gaagTCCCGAGGCTCCCGAGcaggca 146 SV40, Enhancer Region
 200 tctctTCCCGACCGTCCCGAActctctc 227 c-myc, Affe
 2175 tctctTCCCGACCGTCCCGAacctccc 2201 c-myc, Mensch
 2184 cacctTCCCGACCGTCCCGAacctccc 2210 c-myc, Mensch
 2193 caccTCCCGACCGTCCCGATAagcg 2219 c-myc, Mensch
 258 ctcttTCCC GGCTCCCGACTagccc 283 c-myc, Maus
 905 gcctcTCCCC GGCGCGCTCTccctccc 930 c-myc, Huhn
 -251 cactcTCCCGA TCTCCCCCGctctccc -226 β -Aktin Gen, Huhn
 -243 ccactcTCCCGCGCTCCCGAcccccaa -217 β -Aktin Gen, Huhn

Abbildung 10

sitiven Stellen des menschlichen c-myc Gens bestimmt worden. Eine davon, genannt III₁, ist die Homopyrimidin Folge mit dem repetitiven Nonanukleotid, und auch die - allerdings nur grob kartierte - DNase I hypersensitive Stelle II des Hühner c-myc Gens (Schubach & Groudine, 1984) liegt in diesem Bereich.

4.2.6.3.

Die Konsensussequenz der Spl Bindungsstelle

Einige bp 5' zur TATA Box des 1. Promotors der c-myc Gene von Mensch, Affe, Maus und Huhn bzw. zur TATA Box der hauptsächlich verwendeten cap-Stelle des menschlichen N-myc Gens liegt eine Sequenz, die das Hexanukleotid CCGCCC enthält (Abb. 11).

Exakt dieses Hexanukleotid ist in nahezu allen Bindungsstellen des promotorspezifischen Transkriptionsfaktors Spl (Dyan & Tjian, 1983a, 1983b) enthalten (Übersichtsartikel: Kadonaga et al., 1986). Eine Ausnahme bilden 2 der 3 Spl Bindungsstellen im LTR (long terminal repeat) des HTLV-III (Jones et al., 1986). Optimale Bindung, die im entsprechenden Test einen Schutz vor Nuklease Spaltung oder Methylierung liefert, setzt Übereinstimmung mit der Konsensussequenz PQQCCGCCCV (Jones & Tjian, 1985) voraus.

Bei der Wirkung auf die Transkriptionsaktivität kommt es jedoch

zur vorhergehenden Seite

Abbildung 10: Die Polypyrimidin Sequenz im 5'-untranskribierten Bereich von c-myc.

Die Polypyrimidin Sequenz im 5'-untranskribierten Bereich der c-myc Gene von Mensch, Affe, Maus und Huhn im Vergleich mit der Polypyrimidin Sequenz im 5'-untranskribierten Bereich des Hühner β -Aktin Gens (Kost et al. 1983) und der S1 hypersensitiven Stelle (Evans et al., 1984) der SV40 Enhancer Region. Das Sequenzmotiv TCCCCA ist fett gedruckt. Die Pfeile unter den Sequenzen weisen auf die jeweils längsten Repetitionen hin. Unterstrichene Sequenzen sind ebenfalls repetitiv.

weniger auf die Stärke der Bindung als auf den Abstand zur TATA Box an (Gidoni et al., 1985; Jones et al., 1985).

Die potentiellen Spl Bindungsstellen der c-myc Promotoren von Mensch, Affe und Maus und des N-myc Promotors stimmen, bis auf das erste T, mit der geforderten Konsensussequenz überein. Ein T an dieser Stelle hat aber auch die SV40 Spl Box I, die den cap-Stellen der frühen Transkripte am nächsten ist. Eine Mutation dieser Erkennungsstelle reduziert die Aktivität des frühen SV40 Promotors um 80% (Gidoni et al., 1985).

Die Sequenz 209-224 im N-myc Exon 1 ist mit der HSV IE3 Spl Box I (Jones & Tjian, 1985) und 5 benachbarten Nukleotiden (Cordingley et al., 1983) identisch.

295	aaTCTCCGCCCAca	309	c-myc Affe, TATA Box bei Pos. 316
2279	aaTCTCCGCCCAcc	2293	c-myc, Mensch, TATA Box bei Pos. 2300
376	aaTCCCCGCCCAcc	390	c-myc, Maus, TATA Box bei Pos. 397
1122	tgCTCCCGCCCCcg	1136	c-myc, Huhn, TATA Box bei Pos. 1176
24	tcTCCCCGCCCCcc	38	N-myc, Mensch, TATA Box bei Pos. 59
	PQCCCGCCCV		Konsensussequenz der Spl Erkennungsstelle
52	atTCTCCGCCCCat	38	SV40 Spl Box I, TATA Box bei Pos. 21
-72	gcCCCCCGCCCCcccggc	-89	HSV IE3 Spl Box I, TATA Box: -29
207	caCCCCCGCCCCcccggc	224	N-myc, Mensch

Abbildung 11: Potentielle Spl Bindungsstellen in myc Promotor Regionen.

Potentielle Spl Bindungsstellen der c-myc Promotor Regionen von Mensch, Affe, Maus und Huhn und der Promotor Region des menschlichen N-myc Gens im Vergleich mit der Konsensussequenz, der SV40 Spl Box I (Gidoni et al., 1985) und der HSV-I IE3 Spl Box I (Jones & Tjian, 1985). V steht für: C oder A.

4.2.6.4.

Der zum hsp83 HSE homologe Bereich

Der Bereich 2386-2429 des menschlichen c-myc Gens, identisch mit der Sequenz TC7mycl 402-445, zeichnet sich durch Repetitionen aus (Abb. 12A).

Direkt repetiert ist das Pentanukleotid AGGGC. Invers repetiert ist das Pentanukleotid GCCTC bzw. GAGGC, das als Erkennungsstelle für die SV40 und Polyoma T-Antigen Bindung charakterisiert ist (Tjian, 1978; DeLucia et al., 1983; Pomerantz & Hassel, 1984; Wright et al., 1984). Invers repetiert ist auch das Hexanukleotid GAGAAG. Beide inverse Repetitionen zusammen haben große Ähnlichkeit mit dem Hitzeschock Element (HSE) des *Drosophila* hsp83 Gens.

Die Hitzeschock Elemente der hsp Gene (hsp = heat shock protein) sind 5'-untranslatierte Bereiche, die für die Aktivierung der Transkription nach Induktion durch Hitze oder andere Belastungen verantwortlich sind (Übersichtsartikel: Pelham, 1985). Sie enthalten die Stammsequenz C--GAA--TTC--G. Die invers repetierten c-myc Genabschnitte enthalten diese Stammsequenz nicht exakt.

Die Pos. 2418 des menschlichen c-myc Gens wurde von Siebenlist et al. (1984) als DNase I hypersensitive Stelle (genannt III₂) charakterisiert. DNase I hypersensitiv sind auch die Promotorregionen der hsp Gene. Wu (1984) und Parker & Topol (1984) haben nachgewiesen, daß bestimmte Proteine spezifisch an die Hitzeschock Elemente der *Drosophila* Gene hsp70 und hsp83 binden. hsp83 besitzt ein Hitzeschock Element mit relativ langer, inverser Repetition. Das HSE des menschlichen hsp70 Gens ist flankiert von der inversen Repetition der T-Antigen Erkennungsstelle. Die verglichenen Sequenzen des *Drosophila* hsp83 Gens und des c-myc Gens der Maus zeigen das Pentanukleotid GCCTC einmal.

Im c-myc Gen des Huhns ist in diesem Fall kein entsprechendes

Element zu finden, was wegen der großen Sequenzunterschiede verständlich ist. Eine andere Erklärung wäre, daß dieses Element die hauptsächliche Verwendung der cap-Stelle 2 im c-myc Protoonkogen der Säugetierzellen determiniert. Im c-myc Gen des Huhns wird dagegen hauptsächlich die weiter 5' gelegene cap-Stelle benutzt (Linial & Groudine, 1985).

Die distalen Hälften der invers repetierten c-myc Sequenzen und des menschlichen hsp70 HSE sind Teil einer 14/17 bp Homologie (Abb. 12B) mit entgegengesetzter Orientierung. Das Penta-nukleotid GCCTC ist, mit 7 bp Zwischenraum, imperfekt repetiert. Diese Konstellation wäre, bei perfekter GCCTC Repetition und passender (kompatibler) Sequenz des Zwischenraums, eine ideale T-Antigen Bindungsstelle (Jones & Tjian, 1984; Ryder et al., 1985).

Die TC7myc1 Sequenz 389-411 (beim Menschen entsprechend 2373-2395) zeigt zudem Homologie zum Bereich 142-169 des SV40 Enhancers (Abb. 12B).

zur vorhergehenden Seite

Abbildung 12: Ein c-myc Sequenzmotiv mit Homologie zu Hitzeschock Elementen

Teil A zeigt den Vergleich des durch Repetitionen (---▶) ausgezeichneten c-myc Exon 1 Bereichs mit den Hitzeschock Elementen des *Drosophila* hsp83 (Wu, 1984) und des menschlichen hsp70 (Hunt & Morimoto, 1985) Gens. Die Entfernungsangaben zur jeweiligen TATA Box beziehen sich auf die Nukleotide in Höhe des Pfeils (↓). Die Stammsequenz der Hitzeschock Elemente ist fett gedruckt.

Teil B zeigt die Homologie zwischen Sequenzen des c-myc Gens der Primaten, dem menschlichen hsp70 Gen, den SV40 T-Antigenbindungsstellen I und II und der Sequenz 146-169 der SV40 Enhancerregion.

4.2.6.5.

Die vollständig konservierte Sequenz nach der cap-Stelle 2

Vollständig konserviert zwischen Mensch, Affe und Maus sind die ersten 26 Nukleotide nach der c-myc cap-Stelle 2 (Pos. 2490, 507 bzw. 588). Dieser Bereich könnte folgende Sekundärstruktur annehmen und möglicherweise Einfluß auf die Initiation der Transkription, die Prozessierung oder die Translation der mRNA haben:

```

5' ACTCGCTGTAGT
   ||||| | |A
3' GAGCGACCTTA
  
```

4.3.

p9 homologe DNA des TC7 Genoms

4.3.1.

Isolierung

Beim Durchmustern (3.6.5.) von 10 x 45000 Plaques der TC7 Genbank mit dem Klon p9 als Probe fand ich 11 deutliche Signale, die allerdings schwächer waren, als in den Fällen, wo längere, perfekt komplementäre Nukleotidsequenzen miteinander hybridisieren.

Diese 11 rekombinanten Phagen, dazu 3 von ca. 50 noch etwas schwächer hybridisierenden Plaques, habe ich plaquerein isoliert.

Die Charakterisierung der DNA dieser Phagen durch Restriktionsanalyse und Southern Blot Hybridisierung (3.6.4.) gegen p9 (z.B. Abb. 15) zeigte:

Die Gruppe der 11 Phagen besteht aus nur 4 verschiedenen Klonen, die L450, L450i, j und k; L451 und L451i; L452; L453, L453i, j und k genannt wurden. (Phagen mit derselben Zahl im Namen sind identisch.)

Die Inserts der Phagen L450-L453 stellen überlappende Bereiche eines 27kb Genomabschnitts dar; siehe dazu die gemeinsame Restriktionskarte dieser Phagen (Abb. 13).

Die 3 weiteren Phagen, die ich L457, L458 und L459 genannt habe, sind untereinander verschieden; ihre Inserts sind nicht überlappend und stammen nicht von dem durch L450-L453 definierten Genomabschnitt.

Das 6.1kb BamHI Fragment der Phagen L450, L452 und L453, das 25.5kb BamHI Fragment von L451, das 5.9kb BamHI Fragment von L457, das 22kb BamHI Fragment von L458 und eines oder beide 23kb BamHI Fragmente von L459 hybridisieren gegen p9 (Tab. 3).

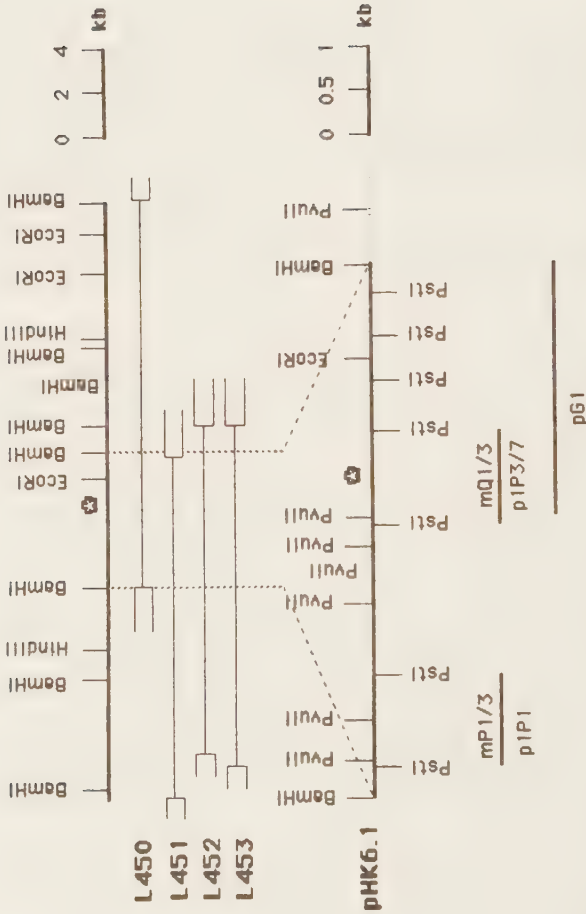


Abbildung 13: Restriktionskarte der rekombinanten Phagen L450-L453 und des Klonas pHK6.1.
Die Inserts der rekombinanten Phagen L450, L451, L452 und L453 sind überlappend und definieren einen 27kb Genomabschnitt der TC7 Zelle. Die Phagenarme (19.6kb, 9.2kb) sind durch längere und kürzere, offene Balken dargestellt. Der Stern zeigt an, welche Fragmente mit p9 hybridisieren. pHK6.1 ist das 6.1kb BamHI Fragment, kloniert in pAT153. Die PvuII Stelle, deren Lage nicht genau bestimmt ist, liegt 320 bzw. 340 bp von den benachbarten PvuII Stellen entfernt. mp1, mp3, plp1, mql, mql, plp3, plp7 und pgl sind Subklone von pHK6.1 in M13 (m) oder pUC12 (p).

L450	L451	L452	L453	L457	L458	L459
19.6	*25.5	19.6	19.6	25	*22	*23(2x)
	10.5	12.5			16	
9.2			9.2	9.2		
6.5					7.0	
*6.1		*6.1	*6.1	*5.9		
	4.5					
	4.2	4.2	4.2			
			3.5			
2.3				2.0		
1.4		1.4	1.4	1.5		
1.2				0.9		

Tabelle 3: Gegen p9 hybridisierende DNA Fragmente
 BamHI Spaltungsmuster der DNA der 7 verschiedenen rekombinanten Phagen. Die mit * markierten Fragmente (Größe in kb) hybridisierten mit p9.

Das 6.1kb BamHI Fragment aus L452 und das 5.9kb BamHI Fragment aus L457 wurden in pAT153 kloniert und pHK6.1 bzw. pHK5.9 genannt.

Die Hybridisierung gespaltener, auf Nitrocellulose Membranfilter übertragener DNA der Subklone pHK6.1 und pHK5.9 gegen p9 zeigte eine - im Vergleich zu den Plasmid Anteilen - sehr schwache Markierung der Inserts (Abb. 14).

Dieser Unterschied in der Stärke der Hybridisierung erschwerte die Eingrenzung der zu p9 homologen Regionen von pHK6.1 und pHK5.9 auf möglichst kleine Fragmente.

Will man die Hybridisierung der Vektoranteile miteinander aus-

schließen, müssen diese aus dem zu untersuchenden Klon und aus dem als markierte Probe dienenden Klon praktisch vollständig entfernt werden. Eine andere Möglichkeit ist, nicht miteinander hybridisierende Vektoren zu verwenden; zum Beispiel pBR322 und M13. Diese Kombination führte neben der Eingrenzung der homologen Bereiche noch zu einem weiteren, völlig unerwarteten Ergebnis (4.3.4.).

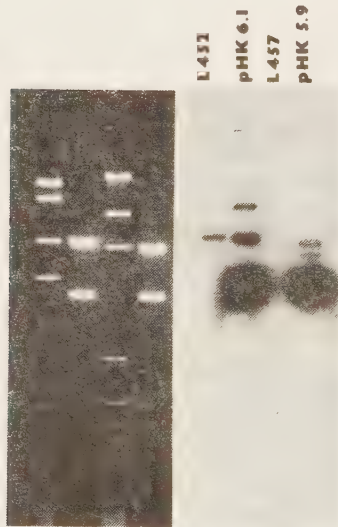


Abbildung 14: Hybridisierung der rekombinanten Phagen L452 und L457 und der Subklone pHK6.1 und pHK5.9 gegen p9
 Links: Das Gel mit der BamHI gespaltenen DNA. Die Fragmentgrößen von L452 (Spur 1) und L457 (Spur 3) sind in Tab. 3 angegeben. Der pAT153 Vektoranteil von pHK6.1 (Spur 2) und pHK5.9 (Spur 4) ist 3.66 kb.
 Rechts: Die Autoradiographie der Southern Blot Hybridisierung mit p9 als Probe.

pHK6.1 hybridisierte sehr stark mit DNA Fragmenten der Phagen L450-L453 aber nur schwach (wie p9) mit DNA aus L457-L459. Dies bestätigte, daß die Inserts der Phagen L450-L453 von der gleichen Stelle des TC7 Genoms stammen, die Inserts der Phagen L457-L459 dagegen aus anderen Bereichen.

4.3.2.

Charakterisierung des L452 Subklons pHK6.1

Hybridisierungsexperimente (z.B. Abb. 15) mit der DNA der Phagen L450-L453 zeigten, daß das 5.0kb BamHI/EcoRI Fragment, das 3.5kb PvuII Fragment, das 2.9kb BamHI/PvuII Fragment und ein 1.1kb PstI Fragment des 27kb Genomabschnitts der TC7 Zelle (Abb. 13) homolog zu p9 sind.

Auf pHK6.1 liegen das 5.0kb BamHI/EcoRI-, das 2.9kb BamHI/PvuII-, und zwei 1.1kb PstI Fragmente (Abb. 13). Das 2.9kb BamHI/PvuII Fragment wurde in pUC12 (BamHI/SmaI) kloniert und pG1 genannt. Die beiden PstI Fragmente wurden in pUC12 (PstI) und in M13mpl0 (PstI) kloniert und erhielten nach Charakterisierung durch PvuII bzw. nach einem Komplementaritätstest (3.6.6.) die Namen plP1 (= pUC12 mit 1.05kb PstI Insert), mP1 und mP3 (M13mpl0 mit 1.05kb PstI Insert in den beiden Orientierungen), plP3 und plP7 (pUC12 mit 1.13kb PstI Insert in den beiden Orientierungen), sowie mQ1 und mQ3 (M13mpl0 mit 1.13kb PstI Insert in den beiden Orientierungen).

Hybridisierungsexperimente mit der DNA der Subklone in M13 zeigten, daß das 1.13kb PstI Fragment Homologie zu p9 hat.

Das Insert von pG1 enthält Schnittstellen für SacI, PstI oder EcoRI. Dadurch konnte ich im Hinblick auf eine Sequenzierung leicht weitere Subklone konstruieren (Abb. 16): mH2, mH5 und pQ1 (340bp SacI Fragment in den beiden entgegengesetzten

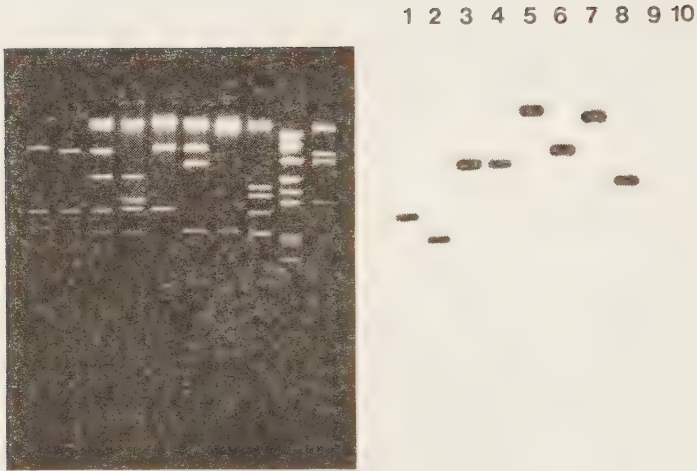


Abbildung 15: Zu p9 homologe DNA Fragmente des 27kb TC7 Genomabschnitts

Links: das Gel, rechts: die Autoradiographie der Southern Blot Hybridisierung mit p9 als Probe. Spur 1-8: L453 DNA, geschnitten mit PvuII (1), BamHI/PvuII (2), BamHI (3); BamHI/HindIII (4), HindIII (5), EcoRI/HindIII (6), EcoRI (7) und BamHI/EcoRI (8); Spur 9: BamHI gespaltene Ad2 DNA und HindIII gespaltene Ad2 DNA (Größenmarker: 14.3, 10.7, 8.1, 6.2, 5.3, 4.7, 3.4, 3.3, 3.2, 2x 2.8, 2.3, 2.1, 1.3 und 1.0kb); Spur 10: SalI gespaltene EMBL3 (Frischauf et al, 1983) Phagen DNA (Kontrolle). Homolog zu p9 sind die L453 DNA Fragmente: 3.5kb PvuII, 2.9kb BamHI/PvuII, 6.1kb BamHI, 29.7kb HindIII, 7.6kb EcoRI/HindIII, 18.4kb EcoRI und 5.0kb BamHI/EcoRI.

Orientierungen in M13mpl9 bzw. in pUC12), mZ1 und pO4 (1.4kb SacI Fragment in M13mpl9 bzw. pUC12), mY2 und pN1 (1.75kb PvuII/EcoRI Fragment in M13mpl1 bzw. pUC12), mS1 und mT1 (1.1kb EcoRI/HindIII Fragment in M13mpl0 bzw. M13mpl1), mU1 und mV1 (0.55kb SacI/SalI Fragment in M13mpl0 bzw. M13mpl1) sowie mG1,

mG2 und pI1 (0.42kb SacI Fragment in beiden Orientierungen in M13mpl9 bzw. in pUC12).

Die Inserts von mQ3, mH5, mZ1 und mY2 haben die gleiche Orientierung, die ich als L (lower) bezeichnet habe, weil, wie es durch die Abbildungen 13, 16 und 17 festgelegt ist, der untere Strang mit dem M13-(+)Strang verbunden ist. mQ1 und mH2 haben die entgegengesetzte U (upper) Orientierung.

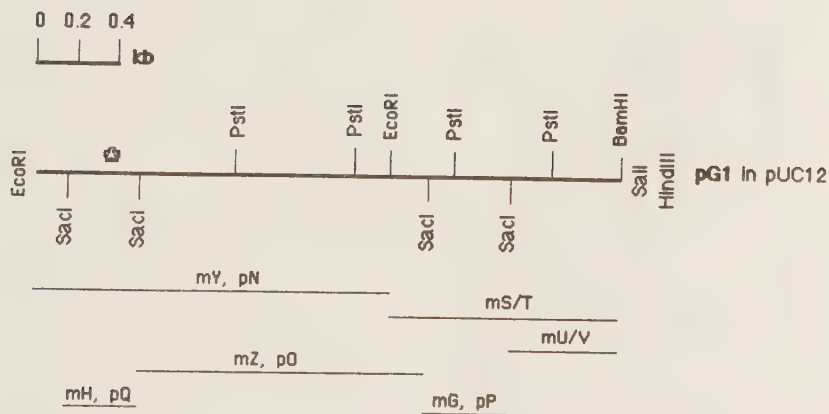


Abbildung 16: Restriktionskarte des Klons pG1

pG1 ist das 2.9kb PvuII/BamHI Fragment des 27kb TC7 Genomabschnitts (Abb. 13), kloniert zwischen die SmaI und die BamHI Stelle von pUC12. Die verbleibenden Schnittstellen des pUC12 Polylinkers sind an den Seiten vermerkt, da sie für einige Subklonierungen verwendet wurden. Größe und Lage der im Text näher bezeichneten Subklone sind dargestellt. Der Stern zeigt den zu p9 homologen Bereich an.

4.3.3.

Sequenzierung des 1.13kb PstI Fragments

Das 1.13kb PstI Fragment, welches mit p9 hybridisierte, wurde nach der Methode von Sanger et al. (1977) sequenziert (3.7.). Es trägt den Namen TC7p9. Die Sequenzierungsstrategie ist in Abb. 17 ausgeführt. Die vollständige Sequenz ist in Abb. 18 dargestellt.

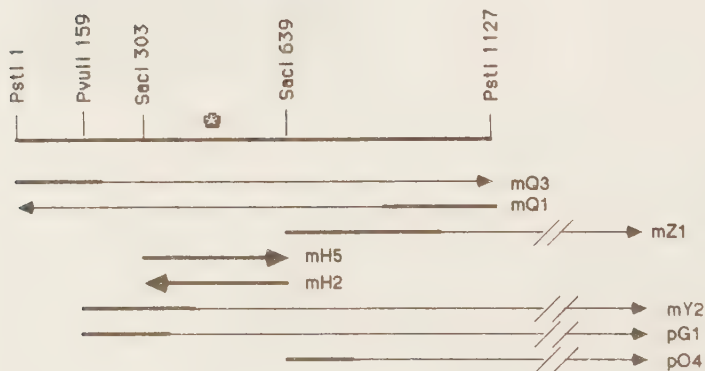


Abbildung 17: Sequenzierung von TC7p9

Die Klone anhand derer das 1.13kb PstI Fragment sequenziert wurde sind aufgeführt. Dicke Linien zeigen die Länge der in Richtung der Pfeile gelesenen Sequenz an. Ein Beispiel: In mH5 (336bp SacI Fragment in M13mpl9) befindet sich der untere (L) Strang. Er diente als Template für die Sequenzierungsreaktion, sodaß die Sequenz des oberen Strangs gelesen werden konnte.

TC7p9

CTGCAGGCCG	GGCACCAGAG	GGGCAGCGAG	TAACCTCAAGG	GGACTCACTC	50
AGCCCACTGT	CTAGCACCAG	GGCCCTGTCT	TGAGTGTGTA	CTTTCCTAAT	100
CGTACTGGGA	GAAACTTAGT	TCTTACAGGA	AGAAACTTAG	TTCCATCACC	150
TCATCTCCCA	GCTGGGGAAA	CTGAGGTCTT	TGAGGTTAAG	ACCCAAGGTC	200
ATATGGAGAG	TTGAGGCAGA	GGGATGCCAG	CCAGAGAGGC	AGAGGGCAGG	250
CCTTGGGGCA	CGGGACGGCC	CCTCCACCCC	AACATCACCT	AGGGACGGGG	300
TCGAGCTCCC	CGCTGCCCCC	CACTCCCCAC	ACCTCACTCA	TCCACAGACA	350
AGCCCCACCT	TCTGTGGACA	CATTTCAACC	TCACCAAATC	AGTAATTAT	400
TGACAGTTGG	TGAAGAGGAG	ATTTCACTTT	TTCCTCCTGT	TAAACTGCTC	450
CTAATTAJTT	CCCTGTCAAC	ACGGCTGGCT	CGCCCGCCCG	CGCCCGCCCG	500
GTTACGCAG	TAATCCCCTG	GCACAGTGCC	TCATTAACCT	TTTGCTCCTC	550
TCCTGCCGCC	TGCCAACAAG	ATGAAGACAA	AAACAGATTG	TTGAACAATG	600
GCTCCGGCTT	GTACGCGCAG	CCCCGTTCCC	CAGCCCCCGA	GCTCCAGGG	650
GAACAGTGGG	GAAAAGAGAG	GGGACTGGGG	TGGGGGAGAG	CCCCAGAGGA	700
ATCAAGGAAG	GGAGAAAGAG	GCACAAAGAA	AGATTGAAGC	AGAAACAGAG	750
GCAGGGACGT	GGGGAGGGGC	CTTGGCAAGG	TAGTGCTTCT	TGCTCCAGAG	800
AGACATGAGG	TTCATTAAAC	AGGATTGGAA	GTGCTCCCAG	GTAGGAAACA	850
CCGGCTGTCA	GAGTTACAGG	GACATTTGCC	ACCACCCGCT	CCAGCCTCTT	900
GTGGTAGCCA	AGCCCATGTT	TCCAGACACT	GGACCCTGCA	TATCTCCCTC	950
AAGGGACCTG	CGGCCCCAGT	GCATGTCTGT	GTTGCGGTGG	GGAGGACAGA	1000
GAGAACCCCA	TGTGCTTGGA	ATCCAGAGCC	CCAATAACCC	AGCGGACCAG	1050
GCCAGAAAGG	CCCACACAGT	CTGGCCTTAC	CACAAGGGAG	GCCAGCACCA	1100
GCCCCACCTA	CAGCGGCCCC	GGCAAGCTGC	AG		1132

Abbildung 18: Nukleotidsequenz TC7p9

TC7p9, die Sequenz des 1.13kb PstI Fragments aus pHK6.1, welches mit p9 hybridisiert. Der U Strang ist dargestellt.

4.3.4

Das p9 Hybridisierungsphänomen

Die pHK6.1 Subklone in M13 zeigten ungewöhnliches Verhalten bei der Hybridisierung mit p9 als Probe: Rekombinante (+)-Strang DNAs der pHK6.1 Subklone in M13 hybridisierten nur mit p9, wenn sie, wie z.B. mH5 und mQ3, den L Strang des 336bp SacI Fragments (TC7p9 639-303) enthalten. Klone mit den dazu komplementären U Strängen hybridisierten nicht (Abb. 19). Dieses Hybridisierungsphänomen trat unabhängig davon auf, ob p9 mit radioaktivem CTP oder GTP markiert wurde.

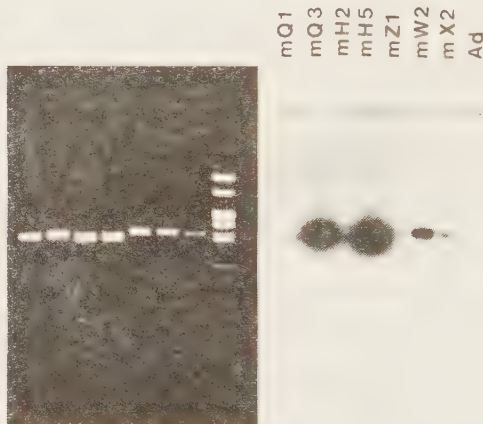


Abbildung 19: Hybridisierung von pHK6.1- und pHK 5.9 Subklonen gegen p9

Links: Das Gel mit der (+)-Strang DNA der rekombinanten M13 Klone mQ1, mQ3, mH2, mH5 und mZ1 (pHK6.1 Subklone), mW2 und mX2 (pHK5.9 Subklone) sowie HindIII gespaltener Ad2 DNA (Größenmarker: 8.1, 5.3, 3.4/3.3/3.2, 2x 2.8, 2.3, 2.1, 1.3 und 1.0kb). Rechts: Die Autoradiographie der Southern Blot Hybridisierung mit p9 als Probe.

Es hybridisierten nur die pHK6.1 Subklone in M13, die den L Strang des 336bp SacI Fragments enthalten. Auch die beiden pHK5.9 Subklone mit komplementären Inserts (mW2, mX2) hybridisierten unterschiedlich stark.

Isoliert man DNA aus M13 infizierten Bakterien (3.3.6.2.), so bekommt man, analog zur Plasmidisolierung, die relativ kleinen Moleküle: doppelsträngige, replikative Formen (RF DNAs) und, in geringem Ausmaß, (-)-Stränge der M13 DNA.

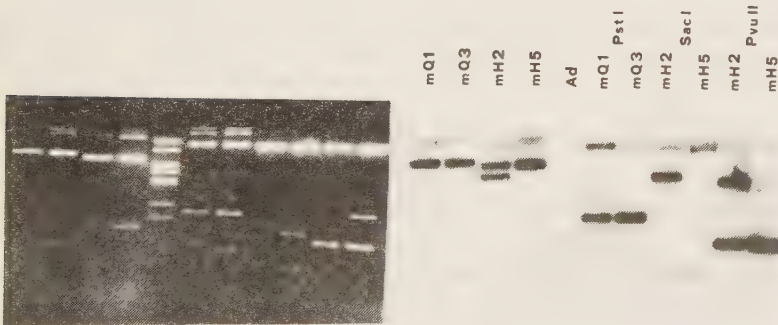


Abbildung 20: Hybridisierung der rekombinanten M13 Klone mQ1/3 und mH2/5 gegen p9

Links: Das Gel zeigt ungespaltene RF DNAs der pHK6.1 Subklone mQ1, mQ3, mH2 und mH5 (Spur 1-4), HindIII gespaltene Ad2 DNA (Spur 5; Fragmentgrößen: 8.1, 5.3, 3.4/3.3/3.2, 2x 2.8, 2.3, 2.1, 1.3, 1.0 und 0.31kb), PstI gespaltene mQ1 und mQ3 DNA (Spur 6 und 7; 7244bp M13mpl0 DNA, 1126bp PstI Insert), SacI gespaltene mH2 und mH5 DNA (Spur 8 und 9; 7250bp M13mpl9 DNA, 336bp SacI Insert) sowie PvuII gespaltene mH2 und mH5 DNA (Spur 10 und 11; das 659bp Fragment besteht aus dem 336bp SacI Insert und flankierenden M13mpl9 DNA Bereichen).

Rechts: Die Autoradiographie der Southern Blot Hybridisierung mit p9 als Probe. Spezifisch markiert wurden die ungespaltenen RFI DNAs, das Insert und 2 (-)-Strang Formen von mQ1, das Insert von mQ3, das Insert und eine (-)-Strang Form von mH2 und das Insert von mH5. Die auf dem Gel unsichtbare (-)-Strang Form der mH2 DNA ergab ein so starkes Signal wie das Insert. Die 336bp SacI Fragmente (Spur 8 und 9) sind nur als schwache Banden erkennbar, weil so kleine Fragmente schlecht an Membranfilter binden.

Doppelsträngige RF DNAs der pHK6.1 Subklone in M13, die das 336bp SacI Fragment enthalten, ließen sich unabhängig von der Orientierung des Inserts durch Hybridisierung mit p9 markieren. Die (-)-Stränge hybridisierten dagegen wiederum nur, wenn sie den L Strang des 336bp SacI Fragments enthielten. Dann aber sehr gut. Denn obwohl in so kleinen Mengen vorhanden, daß sie im gefärbten Gel unsichtbar blieben, ergaben sie ein ungefähr gleich starkes Signal wie die Doppelstränge (Abb. 20). Einzelsträngige, rekombinante M13 DNA ist ein hervorragender Partner für eine Hybridbildung.

Durch diese Ergebnisse und natürlich auch durch die Sequenzierung beider Stränge des 336bp SacI Fragments war zwar schon bewiesen, daß die M13 Klone mit der U Orientierung der Inserts keine Deletionen aufweisen. Dennoch habe ich die Inserts der RF1 DNAs von mH2 und mH5 isoliert und diese mit Fnu4HI, MboII oder AluI gespalten. Die Analyse auf dem Polyacrylamid Gel zeigte identische Fragmente im - aufgrund der Sequenzierung - erwarteten Spaltungsmuster.

Eine Hypothese zur Erklärung dieses p9 Hybridisierungsphänomens habe ich im Abschnitt 4.3.8. aufgestellt.

4.3.5.

Charakterisierung des L457 Subklons pHK5.9

pHK5.9 wurde bisher nur durch Spaltung mit HindIII, PstI oder PvuII charakterisiert. Der zu p9 homologe Bereich wurde auf ein 1.9kb PstI/BamHI Fragment eingegrenzt, in pUC12 kloniert, daraus durch Spaltung mit EcoRI/HindIII wieder isoliert und in M13mpl0 bzw. M13mpl1 eingebracht. Die Klone heißen pH1, mW2 und mX2 (Abb. 21).

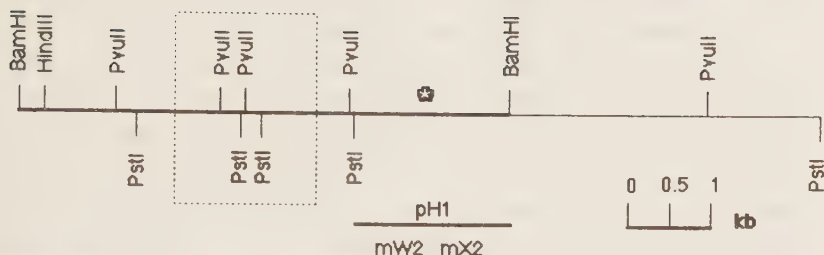


Abbildung 21: Restriktionskarte des Klons pHK5.9

pHK5.9 ist das in pAT153 klonierte 5.9kb BamHI Fragment aus L457, welches gegen p9 hybridisiert. Die Angaben im gestrichelten Kasten bezeichnen nicht die genaue Position der Schnittstellen sondern nur die Größe der PvuII- und der PstI Fragmente dieses Bereichs. Flankierend zum 5.9kb BamHI Fragment sind zwei weitere Schnittstellen angegeben. Der Stern zeigt an, welche Fragmente mit p9 hybridisieren. Die Lage der im Text beschriebenen Subklone pH1, mW2 und mX2 ist ebenfalls angegeben.

4.3.6.

Genomische Analyse der TC7 Zelle bezüglich p9 und pHK6.1

Je 12µg TC7 DNA wurden mit verschiedenen Restriktionsenzymen gespalten, mittels Gelelektrophorese aufgetrennt, auf Membranfilter transferiert und gegen radioaktiv markierte p9 DNA hybridisiert. Diese genomische Analyse mit p9 als Probe, zeigte jedoch keine spezifisch markierten Banden. Wegen der schwachen Hybridisierung von p9 mit den gefundenen rekombinanten Phagen aus der TC7 Genbank kann man Signale auch nur erwarten, wenn der entsprechende hybridisierende Bereich repetitiv wäre. Doch darauf deuten weder die Restriktionskarte des 27kb Genom Abschnitts noch die Anzahl der gefundenen positiven Plaques

hin. Wollte man eine genaue Analyse des TC7 Genoms bezüglich p9 machen, müßte man ein 2-dimensionales Verfahren (z.B.: Tilghman et al., 1977) anwenden.

Vor der Genomanalyse mit pHK6.1 mußte getestet werden, ob dieser Klon eine hoch repetitive Sequenz wie z.B. ein "Alu Element" enthält. Ein entsprechender Test mit radioaktiv markierter, genomischer TC7 DNA als Probe zeigte, daß pHK6.1 und pHK5.9 frei von hoch repetitiven Sequenzen sind. L452 enthält eine allem Anschein nach mittel repetitive Sequenz auf dem 1.65kb PstI- bzw. 2.2kb PvuII Fragment. L457 enthält ein hoch repetitives Element auf seinem 1.25kb PstI- bzw. 1.5kb PvuII Fragment.



Die Genomanalyse der TC7 Zelle mit pHK6.1 als Probe zeigte eine markierte 6.1kb Bande in der Spur mit der BamHI gespaltenen DNA. Ausserdem waren eine ca. 20kb und eine schwache 9.4kb EcoRI Bande, eine 1.7kb und eine 1.1kb PstI Bande (Abb. 22), eine 13.9kb HindIII Bande, eine 3.5kb und eine 1.35kb PvuII Bande sowie eine ca. 5kb und eine 1.4kb SacI Bande zu sehen. Dies steht im Einklang mit der Restriktionskarte des 27kb Genomabschnitts (Abb. 13). Die Intensität der Banden ließ darauf schließen, daß das 6.1kb BamHI Fragment des Klon pHK6.1 im Genom der TC7 Zelle nur einmal vorkommt.

Abbildung 22: Genomische Analyse der TC7 Zelle bezüglich pHK6.1

Folgende vergleichbare Signale fanden sich bei der Analyse der genomischen DNA der menschlichen Tumorzelllinie BL64: eine $\approx 5\text{kb}$ BamHI Bande, eine $\approx 7.5\text{kb}$ - und eine schwächere $\approx 9\text{kb}$ EcoRI Bande und drei PstI Banden ($\approx 2.8\text{kb}$, $\approx 1.7\text{kb}$, $\approx 1.1\text{kb}$).

4.3.7.

Sequenzvergleichsanalyse

4.3.7.1.

Sequenzhomologien zwischen TC7p9 und p9

Der Teil von TC7p9, der mit p9 hybridisiert, wurde experimentell auf das 336bp SacI Fragment und schließlich auf das AluI/SacI Fragment, Pos. 391-639, eingegrenzt.

Der Sequenzvergleich zeigt, daß die größte und auffälligste Homologie zwischen TC7p9 (Abb. 18) und p9 (Anhang 7.1.4.) auf diesem Fragment liegt.

Es sind die nahezu identischen GC₃-Cluster:

p9 199 GCCCGCCCCGCCCGCCCCGCCCGCCCCG 223
TC7p9 478 GCTCGCCCCGCCCGCC GCCCGCCCCG 501
* *

Die beiden Sequenzen stimmen zu 92% überein. Die Wahrscheinlichkeit des zufälligen Auftretens dieser Homologie liegt bei 5.4×10^{-12} (Berechnung siehe Anhang 7.4.).

Insgesamt finden sich auf dem AluI/SacI Fragment von TC7p9 21 Homologien zu p9, die mindestens 12 bp lang sind und mindestens zu 80% übereinstimmen. 20 davon sind Homologien der beiden oberen (U) Stränge, 11 davon wiederum sind Homologien des TC7p9 GC₃-Clusters mit dem p9 GC₃-Cluster (8) und mit dem p9 Bereich 362-395 (3). Ein Teil dieses p9 Bereichs (355-378) zeigt dazu noch 2mal Homologie zum TC7p9 Bereich 549-569.

Daraus 2 Beispiele:

```
p9      362 CGCCTGCCTCGCCCTGCCTCGCCTCCC 388
TC7p9   472 CGGCTGGCTCGCCC GCC CGCCGCC 496
          * * * * *
      81%,  $4.8 \times 10^{-9}$ 
```

```
p9      355 TCTCCGGC GCCTGCC 369
TC7p9   549 TCTCCTGCCGCTGCC 564
          * *
      88%,  $5.1 \times 10^{-7}$ 
```

Wesentlich für die Hybridisierung sind demnach die beiden GC₃-Cluster, die in zahlreichen Variationen zusammenfinden können.

4.3.7.2.

GC₃-Cluster

Ähnliche GC₃-Cluster finden sich in einigen bekannten Sequenzen:

```
84 CCCCCCGCTCCGCCCCCTCCGCCCGTCCGCCCTCCGCCCGCCCCGGCG 35
aus pCaOri5.01, einer SV40 homologen Sequenz des Affen
(Queen et al., 1981)
```

```
106 GCGCCCCGGCCCCGCCCGCCCCGCCGCCG 79
aus pCaOri7.01, einer SV40 homologen Sequenz des Affen
(Queen et al., 1981)
```

```
1389 GCCCGCCTGCCCGCCAGCCC 1370
aus dem Intron 2 des Insulin Gens des Menschen
(Bell et al., 1980)
```

91 GCGCTCGCCCGCCCGCCACC 72

aus der nicht transkribierten 5'-Region eines 5S rRNA
Gens des Hamsters (Hart & Folk, 1982)

403 CCCGCGCGCCCGCCCGCCC 420

aus der nicht transkribierten 5'-Region des Thymidin-
kinase Gens des Huhns (Kwoh & Engler, 1984)

28 GCTGGCTCGCCCGCCCGGGGCGGCC 1

aus dem nicht transkribierten Zwischenraum des rRNA Gen
Clusters des *X.borialis* (Bach et al., 1981)

414 CCCCGGGCCCGCCCGCCCGCCTCCCGCGCGGGGCCCGCCC 454

aus dem inneren transkribierten Zwischenraum - ITS,
internal transcribed spacer - eines rRNA Gens des
X.laevis (Hall & Maden, 1980)

-66 CCCGCCGCCCCCCCCCCCCCGGCCGCCCCGCGCCCGCCCGCCCGGCC -19

aus dem äußeren transkribierten Zwischenraum - ETS,
external transcribed spacer - eines rRNA Gens des
X.laevis (Maden et al., 1982)

17206 CCCCCGCGCCCGCCCGCCG 17188

aus der 4x 500bp Repetition am rechten Ende der
Epstein-Barr Virus DNA (Baer et al., 1984)

516 CGCGCGCCCGCCCGCCCGCCCGCGCGGC 543

aus der nicht transkribierten 5'-Region des IE3 Gens
(IE = immediate early) des Herpes Simplex Virus Typs 2
(HSV-2) (Whitton & Clements, 1984)

162 GCCCTGCCGCCCCGCCCGCCCGCCCGCCCGCC 131

aus der a'-Region der LS Inversion (LS = long short)
des HSV-2 (Davison & Wilkie, 1981)

pCaOri5.01, 7.01 und 9.32 (McCutchan & Singer, 1981; Queen et al., 1981; siehe Anhang 7.1.) sind 3 Klone, ebenfalls aus dem Genom der Affenart *Cercopithecus aethiops*, die aufgrund ihrer Hybridisierung mit dem HindIII C Fragment der SV40 DNA aus einer Genbank isoliert worden sind. Sie zeigen Homologien mit der SV40 Region 5225-105 (Numerierung nach dem BBB-System; siehe Anhang 7.1.3.), einem Teil der Kontrollregion des Virusgenoms.

pCaOri7.01 enthält 2 Promotoren und 2 Bereiche, genannt α und β , die Spl Transkriptionsfaktoren (siehe 4.2.6.3.) binden können (Saffer & Singer, 1984; Gidoni et al., 1984; Dynan et al., 1985).

Die GC₃-Cluster enthalten einige Male die kanonische Stammsequenz (GGGCGG), nicht aber die komplette Konsensussequenz der Spl Erkennungsstellen (Übersichtsartikel: Kadonaga et al., 1986). Identisch mit der HTLV-III Spl Box III (Jones et al., 1986; Anhang 7.2.7.) ist die p9 Sequenz 432-441. Die HTLV-III Spl Boxen I und III bilden jedoch die Ausnahme im Bezug auf die Konsensussequenz.

4.3.7.3.

Gemeinsamkeiten von p9, TC7p9, den pCaOri Klonen und der regulatorischen Region der SV40 DNA

Der Vergleich der Basenzusammensetzung zeigt weitere Gemeinsamkeiten zwischen p9, dem TC7p9 SacI Fragment mit dem GC₃-Cluster, den pCaOri Klonen und der regulatorischen Region der SV40 DNA (Tab. 4):

Hoher G+C Anteil.

Erhöhtes Vorkommen des CG Dinukleotids. [Das CG Dinukleotid, das statistisch einen Anteil von 6.25% ausmachen sollte, hat in der DNA der Vertebraten einen stark verringerten Anteil von nur

2.2% (Nussinov, 1980). Dies hängt mit der spezifischen Methylierung dieses Dinukleotids zusammen. Die Methylierung wiederum hat Bedeutung für die Genexpression.]

Starke Abweichung von der statistischen Verteilung der Basen innerhalb eines Strangs. [Ein C/G Verhältnis von ca. 2:1 und ein Q/P Verhältnis von ca. 3:2, wenn man jeweils den C reicheren Strang betrachtet.]

p9, TC7p9, pCaOri5.01 und pCaOri9.32 haben einige kleinere Repetitionen. Die Repetitionen von pCaOri7.01 haben dagegen ein mit der SV40 Kontrollregion vergleichbares Ausmaß (Tab. 5).

Der Vergleich der Sequenzen von TC7p9 oder p9 mit den pCaOri Klonen und SV40 zeigt ein ganzes Netzwerk von zum Teil recht signifikanten Homologien, in das die GC₃-Cluster, Repetitionen und Spl bindende Regionen in besonderem Maße einbezogen sind. Siehe dazu die Tabellen 6, 7 und 8.

	bp	G+C Gehalt	CG Anteil	C/G Verhältnis	Q/P
p9	529	70.1%	61(11.6%)	2.01	1.99
TC7p9 303-638	336	56.0%	15 (4.5%)	2.19	1.56
pCaOri7.01	436	69.0%	42 (9.7%)	1.89	1.44
pCaOri5.01	227	79.7%	29(12.8%)	1.97	1.74
pCaOri9.32	389	57.3%	44(11.3%)	1.72	1.65
SV40 5225-105	124	60.5%	9 (7.3%)	2.26	1.95
SV40	5243	40.8%	27 (0.5%)	1.06	1.05
TC7p9	1132	57.6%	31 (2.7%)	1.10	0.92
TC7p9 1-160	160	53.8%	3 (1.9%)	1.21	1.05
161-320	160	64.4%	5 (3.1%)	0.81	0.74
321-480	160	47.5%	1 (0.6%)	2.30	1.54
481-640	160	61.3%	14 (8.8%)	1.97	1.46
641-800	160	58.1%	1 (0.6%)	0.36	0.36
801-960	160	53.8%	2 (1.3%)	1.21	1.02
961-1132	172	64.0%	5 (2.9%)	1.15	0.75

Tabelle 4: Vergleich der Basenzusammensetzung

Die unterstrichenen Daten zeigen die Gemeinsamkeiten in der Basenzusammensetzung zwischen p9, TC7p9, den pCaOri Klonen und der regulatorischen Region der SV40 DNA (Sequenzen: Abb. 18 und Anhang 7.1.).

TC7p9 303-638 ist das SacI Fragment mit dem GC₃-Cluster, SV40 5225-105 ist der zu den pCaOri Klonen homologe Bereich. Die Quotienten C/G und Q/P beziehen sich auf den jeweils C reicheren Strang, bei TC7p9 durchwegs auf den oberen (U) Strang.

	Repetitionen		Bemerkungen
TC7p9	110-122	131-143	
	213-223	236-246	
	351-360	1010-1019	
	481-490	492-501	GC ₃ -Cluster
	100-120	121-141	18/21 86%
	105-133	162-191	24/31 77%
p9	199-210	211-222	GC ₃ -Cluster
	299-308	492-501	
	364-373	374-383	
	285-300	378-392	13/16 81%
pCaOri5.01	44-59	60-75	2x 16bp oder 4x 8bp
	44-74	76-108	78%
pCaOri7.01	194-210	211-227	dazu 228-237
	210-221	325-336	dazu 227-237
	80-211	212-343	119/132 90%
	enthält perfekte 56bp, 20bp und 11bp Repetition		
pCaOri9.32	37-48	242-253	
	68-77	78-87	dazu 88-93
	30-54	144-167	80% "3x 30bp"
	30-59	236-264	82% "3x 30bp"
SV40	107-178	179-250	72bp Repetition
	62-82	83-103	21bp Repetition
	5230-5242	1-13	umgekehrt perfekt

Tabelle 5: Repetitionen innerhalb der Sequenzen von TC7p9, p9, pCaOri und SV40

Perfekte Repetitionen sind aufgeführt, wenn sie mindestens 10 bp lang sind. Kommt ein Teil einer perfekten Repetition ein drittes mal vor, ist dies in der Spalte "Bemerkungen" angegeben. Nicht perfekte Repetitionen sind aufgeführt, wenn sie mindestens 15 bp lang und mindestens zu 80% homolog sind. Der Grad der Homologie nicht perfekter Repetitionen ist in der Spalte "Bemerkungen" angegeben.

TC7p9	pCaOri5.01	Homologie
677-690 poly P	81-94 78% Repetition	100%
481-500 GC ₃	81-61 2x 16bp Repetition	86%
306-343 C-reich homolog zu SV40	110-73 78% Repetition	72%

	pCaOri7.01	
670-690 ho. zu pCaOri5.01	203-223 β-Region	76%
756-770	206-219 β-Region	87%
318-343 ho. zu pCaOri5.01 und zu SV40	213-188 β-Region	73%
479-501 GC ₃	108-82 GC ₃	85% 1.8x10 ⁻¹⁰
483-499 GC ₃	351-335 α-Region	88%

	pCaOri9.32	
481-501 GC ₃	177-155	76%

	SV40	
154-181 77% Repetition	144-171 72bp Repetition	79% 2.3x10 ⁻⁹

Tabelle 6: Sequenzhomologien von TC7p9 mit den pCaOri Klonen und der SV40 DNA

Zusätzlich zu den Positionen der Sequenzen sind strukturelle Besonderheiten, weitere Homologien und - soweit möglich - Hinweise auf die Funktion der betreffenden Regionen angegeben. In der dritten Spalte ist der Grad der Übereinstimmung und - für die beiden signifikantesten Homologien - die Wahrscheinlichkeit des Auftretens angegeben.

p9	pCaOri5.01	Homologie
178-232	101-46	71%
GC ₃	2x 16bp Repetition	3.7×10^{-12}
	78% Repetition	
332-359	89-62	75%
P reich	78% Repetition	
homolog zur SV40 21bp	Repetition	

	pCaOri7.01	
188-224	363-328	82%
GC ₃	α -Region	6.0×10^{-14}
199-227	314-341	83%
GC ₃	α -Region, cap-Stelle	1.4×10^{-10}
199-217	98-80	100%
GC ₃	GC ₃	3.6×10^{-12}
373-399	338-309	73%
enthält Repetitionen	α -Region	

	pCaOri9.32	
199-224	60-35	73%
GC ₃	"3x30bp" Repetition	
219-246	275-250	79%
	"3x30bp" Repetition	
359-375	261-245	88%
homolog zu TC7p9	"3x30bp" Repetition	
361-384	91-68	79%
10bp Repetition	10+10+6bp Repetition	

Tabelle 7: Sequenzvergleich von p9 mit den pCaOri Klonen
 Zusätzlich zu den Positionen der Sequenzen sind strukturelle Besonderheiten, weitere Homologien und - soweit möglich - Hinweise auf die Funktion der betreffenden Regionen angegeben. In der dritten Spalte ist der Grad der Übereinstimmung und - für die 4 signifikantesten Homologien - die Wahrscheinlichkeit des Auftretens angegeben.

A	p9	SV40		Homologie
	7-27	5235-11		81%
	4.5S RNA Gen	13bp inverse Repetition		
	341-370	66-35		75%
	homolog zu pCaOri5.01	21bp Repetition		
	385-405	167-147		86%
	homolog zu TC7p9	72bp Repetition		
	462-483	5235-11		82%
		13bp inverse Repetition		
B	231-248	254-266	260-271	311-327
				341-370
				385-405
				462-483
	14/18	12/14	12/13	14/17
				24/32
	1-5227	211-199	190-178	7-5236
				66-35
				239-219
				5235-11
	ORI	72bp R.	72bp R.	ORI
				21bp R.
				72bp R.
				ORI

Tabelle 8: Sequenzhomologien zwischen p9 und SV40

A: Die größten Homologien zwischen p9 und dem regulatorischen Bereich der SV40 DNA. Zusätzlich zu den Positionen der Sequenzen sind weitere Homologien und Hinweise auf die Funktionen der betreffenden Regionen angegeben. B: Homologie zwischen der p9 Region 231-483 und Sequenzen aus dem regulatorischen Bereich der SV40 DNA. Zeile 2 gibt den Grad der Übereinstimmung, Zeile 4 die Funktionen der zu p9 homologen SV40 Regionen an. ORI, 13bp inverse Repetition am Ursprungspunkt der Replikation; 72bp R., 72bp Repetition des Enhancers; 21bp R., 21bp Repetition mit Spl Bindungsstellen.

4.3.7.4.

Charakterisierung der TC7p9 Sequenz

(Siehe dazu Abb. 23.)

Die TC7p9 Sequenz 100-250 enthält alle größeren Repetitionen. Sie zeigt ausgeprägte Homologie zu dem Teil der SV40 72bp Repetition (Anhang 7.1.3.) mit der Enhancer Stammsequenz 168 CTTTCCACACC 178 und der CTGGGGA Repetition (Anhang 7.2.). Die perfekte 11bp Repetition (213-223; 236-246) trägt das Pentanukleotid GAGGC der SV40 T-Antigen Bindungsstellen I und II (siehe auch 4.2.6.4.). 9 der 11 bp stimmen mit der SV40 Sequenz 5232-5242 überein. Im Zwischenraum der 11bp Repetition liegt eine direkte 4bp Repetition.

Charakteristisch für den TC7p9 Bereich 310-440 sind die Homopurin/Homopyrimidin Sequenzen. Homopolymere DNA kommt oft in regulatorischen Genregionen vor; sie hat eine besondere Struktur und kann durch S1 Nuklease gespalten werden (S1 hypersensitive Stellen, SHS: siehe auch 4.2.6.2.). 8bp der TC7p9 Sequenz 409-417 stimmen mit der SHS des Hühner β -Globin Gens (siehe Anhang 7.2.9.) überein. Die TC7p9 Sequenz 305-334 mit der Repetition CTCCCC zeigt große Übereinstimmung mit der SV40 Enhancer Stammsequenz und der SV40 CTGGGGA Repetition. TC7p9 470-650 ist der Bereich mit der größten Homologie zu p9 (HTC). Wieder finden sich lange Homopurin/Homopyrimidinsequenzen. Bemerkenswert ist auch, daß direkt neben dem GC₃-Cluster und 10 bp weiter, 3' zur Sequenz CCCGC, je eine Folge von 6 alternierenden Purin- und Pyrimidinnukleotiden liegt. Die längste alternierende P/Q Folge von TC7p9 ist allerdings die Sequenz 610-619. Unmittelbar darauf folgt GC₃.

```

          90                      110                      130
TC7p9.....GACTTTCCTAATCGTACTGGGAGAACTTAGTTCTTACAGGA
          166GACTTTCACACACC178          -----13----->
                  ** *          -----77%-----
          -----86%----->

          150                      170                      190
AGAAACTTAGTTCCATCACCTCATCTCCCAGCTGGGGAAACTGAGGTCTTTGAGGTTAAG
-----13----->          144TCTGCCTGCTGGGGAGCCTGGGGACTTT171
          167ACTTTCACACACC179          * *          ** * *
-->          * * *          -----77%-----
-86%----->

          210                      230                      250
ACCCAAGGTCATATGGAGAGTTTGAGGCAGAGGGATGCCAGCCAGAGAGGCAGAGGGCAGG
          -----11----->-----11----->
          5232GAGGCCGAGGC5242          5232GAGGCCGAGGC5242
          * *          * *
          150AGGCAGAAAGTATGCI37
                  * *

          270                      290                      310
CCTTGCGGGCAGCGGGACGGCCCCCTCCACCCCAACATCACCTAGGGACGGGGTCGAGCTCCC
----->          <-----
          168CTTTCACACACC178          160GCTCCC
                  ** *

          330                      350                      370
CGCTGCCCCCCTCCACACCTCACTCATCCACAGACAAGCCCCACCTTCTGTGGACA
          ----->
          C154          167ACTTTCACACACC178          176TGTGGA171
                  **
          166CCCCACTCCCCAC151
          <<>>          <>
          .82*Q.

          390                      410                      430
CATTTCACCTCACCAAATCAGCTAATTATTGACAGTTGGTGAAGAGGAGATTTCACTTT
          -----7----->          12/13P          -----
          117CTAATT122

          450                      470                      490
TTCCTCCTGTTAAACTGCTCCTAATTAGTTCCCTGTCAACACGGCTGGCTCGCCCGCCCG
-10/12->          -----7----->          GC3-
          12Q          122AATTAGT116
          169AGTCCCCAGGCTCCCCAGCAGGC147
                  * * * * *

```

weiter nächste Seite

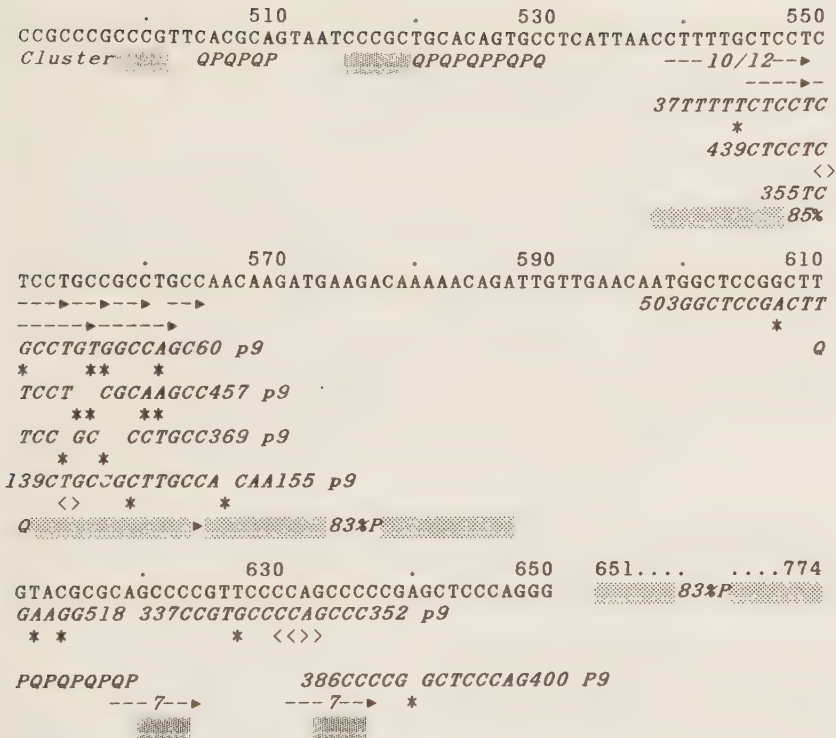


Abb. 23: Homologie von TC7p9 mit p9 und SV40

Die TC7p9 Sequenz 89-650 ist gezeigt. Darunter stehen die Homologien zu p9 (*kursiv*) und zur regulatorischen Region der SV40 DNA (normaler Schrifttyp). Ein Stern * bezeichnet eine verschiedene Base, ein Paar eckiger Klammern <> ein zusätzliches Nukleotid in der homologen Sequenz. Pfeile weisen auf Repetitionen (siehe auch Tab. 5), Balken auf homopolymere Regionen  oder auf das GC₃ Sequenzmotiv  der TC7p9 Sequenz hin.

Die beiden größten Homologien zwischen TC7p9 und p9 (siehe 4.3.7.1.) sind in diesem Diagramm nicht enthalten.

4.3.7.5.

Charakterisierung der p9 (HTC) Sequenz

(Siehe dazu Abb. 24.)

Der GC₃-Cluster von p9 wird flankiert von 6 (Pos. 193) bzw. 5 (Pos. 222) alternierenden P/Q. Einige bp weiter (Pos. 245) liegen 7 + 6 alternierende P/Q. Dieser Bereich ist der SV40 "SphI Region" (Pos. 126-142) sehr ähnlich (siehe auch 4.3.7.6.).

Auffallend an der p9 Homopyrimidin Sequenz 271-308 ist die Wiederholung GQ₈. Wegen dieser Regelmäßigkeit, dem Fehlen von A (Wahrscheinlichkeit 1.8×10^{-5}) und 5 + 5 alternierenden P/Q ähnelt dieser Bereich den GC₃-Clustern. 8 bp weiter (Pos. 316) liegt die Sequenz (GC₃)₂.

Eine weitere Region mit alternierenden P/Q erstreckt sich von 421-463. Zwischen den zuletzt erwähnten alternierenden P/Q Folgen gibt es folgende Homologien:

```

          450 CGCAAGCCGCGCAGCC 464
418 CCGATGCGCGCGCAAGC CACGCC 439
294 CCG CGTCCCGCGC 306
456 CCG CACGCCG 465

```

Die p9 Region 337-405 besteht zu 73% aus Pyrimidinnukleotiden (Q). Sie zeigt 4 größere Homologien mit TC7p9 und 2 größere Homologien mit der SV40 DNA. 5 der 6 längeren, nicht perfekten Repetitionen beziehen sich auf diese Region. Sie trägt eine direkte, perfekte 10bp-, eine 6bp- und eine 8 + 8 + 5bp Repetition. Letztgenannte enthält 3x das Pentanukleotid GCCTC der SV40 T-Antigen Bindungsstellen.

Die SV40 Region 147-167, die zu 86% mit p9 384-405 übereinstimmt, enthält eine S1 hypersensitive Stelle (Evans et al., 1984) und trägt die Repetition CTGGGGA, die wesentlich zur Enhancer Wirkung beiträgt (Zenke et al., 1986).

Als Enhancer Stammsequenz (Weiher et al., 1983) kommen die p9

Sequenzen 146-156 und 429-440 in Frage. Dies zeigt der Vergleich mit SV40 Enhancer Stammsequenzen (siehe Anhang 7.2.2.):

```
146 CTTGCCACAAC 156 p9
429 CAAGCCACGCC 439 p9
168 CTTTCCACACC 178 SV40
    CTTACCACA      MSV
    CTTGCCACA      IgH3
    CAAACCAGTCC    BKV
```

[MSV = Mouse Sarcoma Virus = Murine Sarcoma Virus = Moloney Murine Sarcoma Virus (Dhar et al., 1980); BKV ist ein menschliches Papova Virus (Seif et al., 1979)]

Der in Abb. 24 dargestellte p9 Bereich 95-529 enthält 9x die Stammsequenz der T-Antigen Bindungsstellen und zwar immer in der Orientierung GCCTC. Die beiden ersten Pentanukleotide (Pos. 106 und 121) sind durch 10 bp voneinander getrennt. Die beiden nächsten liegen je ca. 15 bp vom GC₃-Cluster entfernt. Drei weitere (Pos. 367, 377 und 382) sind durch 0, 5 oder 10 bp von einander getrennt.

Die nicht perfekte, inverse Repetition (460-482) schließlich ist der inversen 13bp Repetition der SV40 DNA sehr ähnlich, die eine wesentliche Rolle in der Replikation des Virus Genoms spielt (Shortle & Nathans, 1979; Margolskee & Nathans, 1984):

```
p9      460(5') ACGCCGAGGTT
          || ||||| C
          482(3') TGTGGCTCCGG

SV40 5235(5') AGGCCGAGGC
          |||||G
          10(3') TCCGGCTCCG
```

[illegible]

weiter nächste Seite

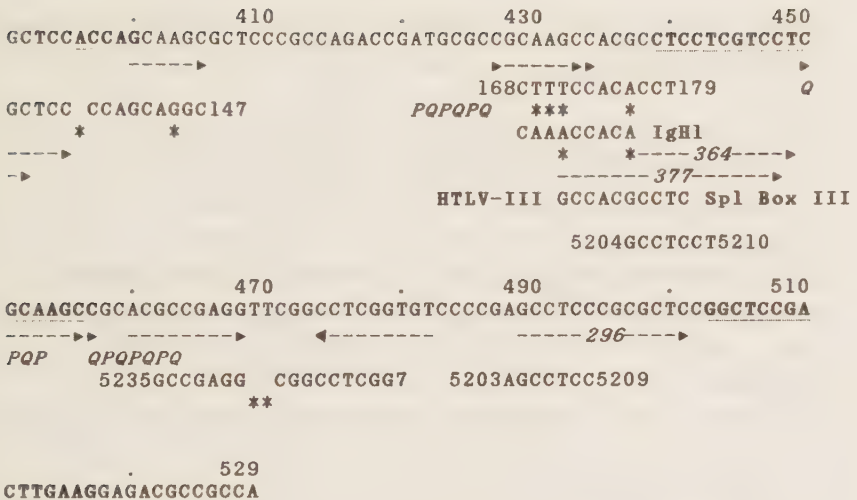


Abbildung 24: Sequenzhomologie von p9 mit TC7p9 und SV40

Die p9 Sequenz 95-529 ist gezeigt. TC7p9 homologe Bereiche (siehe 4.3.7.1. und Abb. 23) sind fett gedruckt und unterstrichen. Unter der p9 Sequenz stehen die Homologien zur regulatorischen Region der SV40 DNA (Anhang 7.1.3.), zu Enhancer Stammsequenzen (Ad5, IgH, MSV), zu SI hypersensitiven Stellen (SHS) und zur HTLV-III Spl Box III (siehe Anhang 7.2.). Ein Stern * bedeutet eine abweichende Base, ein Paar eckiger Klammern < > ein zusätzliches Nukleotid. Repetitionen sind durch Pfeile gekennzeichnet. Bei weiter auseinanderliegenden Repetitionen ist durch kursiv geschriebene Zahlen angegeben, wo die Sequenz erneut vorkommt. Auf Folgen von alternierenden Purin- und Pyrimidinnukleotiden PQ, den pyrimidinreichen Abschnitt 271-308 und auf das Motiv GC₃ der p9 Sequenz ist ebenfalls hingewiesen.

4.3.7.6.

Sequenzen mit alternierenden Purin-/Pyrimidinnukleotiden

Alternierende P/Q Nukleotidfolgen können Z-DNA Struktur (Übersichtsartikel: Rich et al., 1984) annehmen (Wang et al., 1979) und führen in A-, besonders aber in B-DNA zu Strukturabweichungen, um die sterische Hinderung zwischen benachbarten Purinen auf verschiedenen Strängen auszugleichen (Calladine, 1982). Eine bestimmte Anordnung von 2 längeren, alternierenden P/Q Folgen findet sich in den regulatorischen Regionen zahlreicher Virus Genome (Nordheim & Rich, 1983).

Die SV40 DNA enthält überhaupt nur 4 Nukleotidfolgen mit mehr als 7 alternierenden P/Q (siehe Anhang 7.3.). 3 davon sind Bestandteil des Enhancers und sind für dessen Wirkung wichtig (SphI Motiv; Zenke et al., 1986). Auch bestimmte zelluläre Transkriptionskontrollbereiche enthalten SphI Motive (Abb. 25 und Anhang 7.2.3.). Die SV40 Sequenz 126-142 bzw. 198-214 ("SphI Region") enthält 8 + 7 alternierende P/Q und eine SphI Schnittstelle. An die Enhancerregion überspiralisierter SV40 DNA binden Z-DNA spezifische Antikörper (Nordheim & Rich, 1983) und Z-DNA erkennende "PZ-Proteine" (Azorin & Rich, 1985). Die Bindung dieser Proteine zeigt sich durch eine verminderte Spaltung der DNA durch SphI.

zur nachfolgenden Seite

Abbildung 25: Sequenzen mit alternierenden Purin-/Pyrimidinnukleotiden in p9, TC7p9, pCaOri7.01 und in Transkriptionskontrollelementen

A: Der GC₃-Cluster von pCaOri7.01 mit flankierenden Sequenzen.
 B: Vergleich der P/Q Sequenzen von p9, TC7p9 und pCaOri7.01 mit Enhancerelementen (SphI Motiven) von SV40, LPV (Mosthaf et al., 1985), Polyoma (Sequenz: Griffin et al., 1980), Igh (Banerji et al., 1983) und von den Genen der kleinen nuklearen RNAs U1 und U2 (Krol et al., 1985) sowie mit Promotorelementen von V_H (Grosschedl & Baltimore, 1985) und V_K (Seidmann et al., 1979).

C: Das GC₃-Cluster flankierende Sequenzelement TGCGC.

A

pCaOri7.01

110 AACTGCGCCCCGGCCCCGCCGCCGCCGCCGAGCCAGGCCGTGCATGCCCG 58

B

123 GAGATGCATGCTTTGCATACTTCTGCCT 150 SV40 (= 195-222)

245 CCGCACGCT TGCACAA 260 p9

503 TCACGCAGT Na TGCACAG TGCCT 531 TC7p9

934 CCCTGCATAT_{N25} GTGCATGT CTGTGT 982 TC7p9

125 gaTGCGATGcttTGCATActtt 144 SV40

70 cgTGCGATGcc 61 pCaOri7.01

200 gcTGCGACGgg 191 p9

203 ttTGCGAGactt GCAGActt 221 LPV, 106/63bp Repetition

5193 ttTGCAAGag 5202 Polyoma Enhancer

167 ttTGCAATtct 177 Igh Enhancer

-51 ttTGCAActc -61 Vh Promotor

-65 ttTGCAAccc -55 Vh Promotor

ttTPCATP Enhancerelement der U1, U2 snRNA Gene

C

109 acTGCGCcc 101 pCaOri7.01

420 gaTGCGCcg 428 p9

621 gcTGCGCgtaca 610 TC7p9

Charakteristische Bereiche der p9 und der TC7p9 DNA sind die GC₃-Cluster. Diese sind mit alternierenden P/Q Nukleotidfolgen verknüpft (Abb. 23 + 24; 4.3.7.4., 4.3.7.5.). Flankiert von Sequenzen mit alternierenden P/Q ist auch der GC₃-Cluster von pCaOri7.01 (Abb. 25A). Fast alle dieser mit den GC₃ Motiven von p9, TC7p9 und pCaOri7.01 verknüpften Sequenzen enthalten das Hexanukleotid TGCAQP (Abb. 25B) oder das Pentanukleotid TGCGC (Abb. 25C). Auf die Homologie zwischen p9 und der SV40 SphI Region wurde bereits hingewiesen, und auch die TC7p9 Sequenzen 503-531 und 934-982 zeigen dazu gewisse Übereinstimmung (Abb. 25B).

Der TC7p9 Bereich 937-961 zeigt eine Homologie zu einem V_K-Promotor (Seidmann et al., 1979), die über die alternierende Sequenz noch hinausgeht:

--> -----> <----- <---

937 TGCATATCTCCCTCAAGGGACCTGC 961 TC7p9

-63 TGCATA CCCTCA CTGC -48 bezogen auf die V_K cap-Stelle

Der Sequenzvergleich (Abb. 25; siehe auch Anhang 7.2.3. und Falkner & Zachau, 1984; Parslow et al., 1984), DNA-Protein Bindungsexperimente (Azorin & Rich, 1985; Singh et al., 1986) und die Mutationsanalyse von Zenke et al. (1986) lassen stark vermuten, daß das Sequenzmotiv TGCAQP ein wichtiger Bestandteil bestimmter Transkriptionskontrollelemente ist.

Daß die Verknüpfung der GC₃-Cluster mit alternierenden P/Q Sequenzen und speziell mit dem Motiv TGCAQP außergewöhnlich ist, zeigt folgende Betrachtung:

Eine Folge von 6 alternierenden P/Q sollte alle 32 bp, eine von 7 alle 64 bp auftreten, usw. ($P(m) = 2 \times 0.5^m$). Vorausgesetzt ist ein P/Q Verhältnis von 1:1, sonst ist der statistische Erwartungswert entsprechend reduziert.

Tatsächlich sind aber Sequenzen mit alternierenden P/Q in natürlicher DNA mehr oder weniger stark reduziert (siehe Anhang

7.3.), z.B. in eukaryotischen Zellen im Mittel um 50%. Besonders selten sind sie im SV40 Genom.

Eher erhöht ist das Auftreten alternierender P/Q in nicht transkribierten Bereichen einiger Virusgene (Tab. A2 im Anhang 7.3.). Dies könnte funktionelle Bedeutung haben.

Die Sequenzen TGCAQP oder TGCGC sollten, wenn man beide Stränge einer dsDNA in Betracht zieht, bei gleicher Häufigkeit der 4 Basen (je 0.25) alle 512 bp auftreten. Das ergäbe Erwartungswerte von 2.2 für TC7p9 und 1.0 für p9. Berücksichtigt man die spezielle Basenzusammensetzung und die reduzierte Häufigkeit von alternierenden P/Q Folgen, ergibt sich für TC7p9 und p9:

$P(\text{TGCAQP})$ in TC7p9; 1132bp =

$$2 \times 0.18 \times 0.27 \times 0.30 \times 0.25 \times 0.48 \times 0.52 \times 1100 = 2.0$$

Da Hexanukleotide mit alternierenden P/Q in TC7p9 jedoch um 50% reduziert sind (Tab. A1 im Anhang 7.3.), erwartet man die Sequenz einmal. Sie kommt aber 3mal vor, und zwar 3mal im oberen Strang, auf den Positionen 520 (19 bp vom GC₃-Cluster entfernt), 937 und 970.

$P(\text{TGCATP})$ in TC7p9 =

$$0.18 \times 0.27 \times 0.30 \times 0.25 \times 0.18 \times 0.52 \times 1100 \times 0.5 + \\ 0.25 \times 0.30 \times 0.27 \times 0.18 \times 0.25 \times 0.48 \times 1100 \times 0.5 = \\ 0.19 + 0.24 = 0.43$$

Der Erwartungswert der Sequenz TGCATP (Repetition des SV40 SphI Motivs) beträgt demnach 0.19 für den oberen (U) und 0.24 für den unteren (L) Strang. Sie kommt aber 2 mal im U Strang (Pos. 937 und 970) innerhalb von nur 38 bp vor.

$P(\text{TGCAQP})$ in p9; 529bp, 25% reduzierte Häufigkeit von Hexameren mit alternierenden P/Q (Tab. A1 im Anhang 7.3.) =

$$2 \times 0.20 \times 0.23 \times 0.47 \times 0.10 \times 0.33 \times 0.66 \times 530 \times 0.75 = 0.37$$

Die Sequenz kommt aber 2mal vor: direkt neben (Pos. 198) und 30

bp entfernt (Pos. 254) vom GC₃-Cluster.

Das heißt, die Sequenz TGCAQP tritt in TC7p9 (3 : 2.2 : 1.0) und p9 (2 : 1.0 : 0.37) häufiger auf, als man es statistisch erwarten würde. Dies wird besonders deutlich, wenn man die spezifische Basenzusammensetzung von TC7p9 bzw. p9 und die reduzierte Häufigkeit von beliebigen Hexameren mit alternierenden P/Q, wie in obiger Berechnung geschehen, berücksichtigt.

4.3.7.7.

Offene Leserahmen von TC7p9

Betrachtet man TC7p9 im Hinblick auf mögliche kodierende Sequenzen (Abb. 26) sieht man, daß die linke Hälfte voll von Stoppkodons ist, die rechte aber einige, bis zu 483 bp lange offene Leserahmen hat. Beispielsweise könnte der obere (U) Strang im Leserahmen I, beginnend mit ATG 598, gespleißt von Pos. 780 bis 869, für 148 Aminosäuren kodieren. TATA Boxen und Spleiß Konsensussequenzen (Anhang 7.2.8.) wären vorhanden. Für die Konstellation mit 2 nahe beieinander liegenden Initiationskodons, von denen das zweite verwendet wird oder, wegen eines dazwischen liegenden Stoppkodons, nur das zweite verwendet werden kann, gibt es Beispiele (Kozak, 1984).



Abbildung 26: Leserahmen in der Sequenz TC7p9

Die Sequenz TC7p9 ist dargestellt mit den Schnittstellen für PstI und SacI, 2 TATA Box homologen Sequenzen, der Lage des GC3-Clusters und den 6 Leserahmen. Striche auf und unter den durch Linien symbolisierten Leserahmen bezeichnen ATG- bzw. Stoppkodons. Die Leserahmen sind mit U/LRF I-III (upper/lower reading frame) bezeichnet, der Pfeil gibt die Leserichtung an. URF I-III beginnen bei TC7p9 Positionen 1, 2 und 3. LRF I-III bei 1132, 1131 und 1130. Das im Text erwähnte potentielle Intron des URF I ist durch ein Dreieck angegeben.

TC7p9	Ausmaß/ Wahrschein- lichkeit	homologe Sequenz
U 906-937	82%	C α 1 immunoglobulin Genregion des Menschen, 120 bp 5' zum Exon 1 (Ellison et al., 1982)
U 888-939	74%	Der entsprechende Bereich 5' zum Exon 1 der Ig γ 2 (=Ig γ 3; =Ig γ 4) Genregion des Menschen (Takahashi et al., 1982)
U 860-892	79%	C κ 1 immunoglobulin Genregion des Menschen, 3'-untranslatierter Bereich (Hieter et al., 1980)
U 361-386	85%	C κ 2 immunoglobulin Genregion des Menschen, kodierende Region (Bentley & Rabbits, 1983)
U 602-671	65%	c-myc Gen der Maus, 143 bp 5' zur TATA Box 1 (Bernard et al., 1983)
U 706-720	100%	(Pos. 252-323; siehe Abb. 7+8: 4.2.5.)
U 544-562	89%	c-myc Gen der Maus, 47 bp 5' zur TATA Box 1 (Pos. 348-362)
U 625-654	83%	c-myc Gen der Maus, Exon 1 (untranslatiert) (Pos. 474-492)
L 732-870	58%	Präprosomatostatin Gen der Ratte, Intron 1 (Montminy et al., 1984)
L 472-495	88%	Skelettmuskel Aktin Gen der Ratte, Intron 1, innerhalb der 5'-untranslatierten Region (Zakut et al., 1982)
U 255-290	72%	5S rRNA Gen Cluster des Syrischen Hamsters, nicht transkribierter Zwischenraum (Hart & Folk, 1982)
U 293-328	75%	Cytoplasma β -Aktin Gen des Huhns, 5'-untranskribierte Region (Kost et al., 1983)
U 306-333	79%	Cytoplasma β -Aktin Gen des Huhns, Intron 1, innerhalb der 5'-untranslatierten Region
U 759-793	69%	Cytoplasma β -Aktin Gen des Huhns, 5'-untranskribierte Region
	3.6x10 ⁻⁸	Cytoplasma β -Aktin Gen des Huhns, Intron 1, innerhalb der 5'-untranslatierten Region

weiter nächste Seite

U 483-500	94%	Gen für die cytoplasmatische Thymidinkinase des Huhns, 360 bp 5' zur cap-Stelle (Kwoh & Engler, 1984)
L 663-745	7.0x10 ⁻¹⁰	Y Gen aus der Ovalbumin Genfamilie des Huhns, 5'-untranslatierte Region mit GAAG Repetition (Heilig et al., 1982)
L 473-500	61%	rRNA Gen Cluster des <i>X. laevis</i> , nicht transkribierter Zwischenraum (Bach et al., 1981)
L 535-575	86%	EV-2 (Avian Endogenes Retrovirus) LTR flankierende DNA im Genom eines Huhns, 54 bp 5' zur U3 Region (Scholl et al., 1983)
U 308-328	3.1x10 ⁻¹⁰	Epstein-Barr Virus (EBV), 4x 500bp Repetition (Baer et al., 1984)
U 308-330	86%	EBV, BDRF1 (BamHI d fragment rightward reading frame 1)
U 659-684	5.9x10 ⁻⁹	EBV, Region zwischen BBRF3 und BBLF1
U1041-1057	81%	EBV, BPLF1
L 295-327	2.3x10 ⁻⁹	EBV, 4x 500bp Repetition
L 479-501	80%	EBV, 4x 500bp Repetition
L 477-501	3.0x10 ⁻¹⁰	Herpes Simplex Virus Typ 1 (HSV-1), IE3 Gen, 5'-untranskribierte Region (Cordingley et al., 1983) mit Spl Bindungsstelle (Jones & Tjian, 1985)
L 478-508	87%	HSV-2, a' Repetition der LS Inversionsregion (Davison & Wilkie, 1981)
	5.1x10 ⁻¹⁰	
	84%	
	3.4x10 ⁻⁹	
	81%	
	3.5x10 ⁻¹⁰	

Dazu kommt noch die in Tabelle 6 aufgeführte Homologie zu pCaOri7.01.

Tabelle 9: Homologien von TC7p9 mit bekannten Sequenzen (zu Kapitel 4.3.7.8.)

In Spalte 1 ist der Strang und der Sequenzabschnitt von TC7p9 angegeben, der die Homologie zu der in Spalte 3 angegebenen Sequenz hat. Spalte 2 zeigt das Ausmaß der Homologie und die Wahrscheinlichkeit eines zufälligen Auftretens. Es wurden nur signifikante Homologien berücksichtigt. (Zur Orientierung: Sequenzen mit einer Wahrscheinlichkeit eines zufälligen Auftretens von $\leq 2.0 \times 10^{-10}$ sollten im Säugetier Genom statistisch nur einmal vorkommen.)

p9	Ausmaß/ Wahrschein- lichkeit	homologe Sequenz
L 258-427	59%	zeta-Globin Gen des Menschen, Intron 1; (Proudfoot et al., 1982)
L 481-508	4.2x10 ⁻⁹ 82%	
L 204-254	2.2x10 ⁻¹⁰ 75%	Dihydrofolatreduktase Gen des Menschen, 14 bp 5' zur TATA Box, 41 bp 5' zur cap-Stelle (Chen et al., 1984), beinhaltet 2 Spl Konsensussequenzen
L 199-229	1.7x10 ⁻¹⁰ 82%	
U 181-241	1.9x10 ⁻¹⁰ 66%	Insulin Gen des Menschen, Intron 2; (Bell et al., 1980)
U 199-221	4.3x10 ⁻⁹ 91%	
U 290-328	2.7x10 ⁻¹¹ 80%	<i>X. laevis</i> rRNA Gen Cluster, interner transkribierter Zwischenraum 1 (ITS1); (Hall & Maden, 1980)
U 404-465	1.4x10 ⁻⁷ 65%	
L 359-414	6.7x10 ⁻⁹ 77%	Herpes Simplex Virus Typ 2 (HSV-2), ca 200 bp 5' zur cap-Stelle des IE3 Gens; (Whitton & Clements, 1984)
U 180-224	1.8x10 ⁻¹² 76%	
L 177-223	8.2x10 ⁻¹³ 73%	<i>X. laevis</i> rRNA Gen Cluster, externer transkribierter Zwischenraum (ETS) 67 bp 5' zur 18S RNA kodierenden Sequenz, 100% G+C; (Maden et al., 1982)
L 202-224	4.0x10 ⁻¹¹ 91%	
	2.7x10 ⁻¹¹	

weiter nächste Seite

L 192-232	76%	HSV-2, a'-Repetition der LS Inversionsregion; (Davison & Wilkie, 1981)
L 201-224	6.7x10 ⁻¹¹ 88%	- " -
U 192-214	4.3x10 ⁻¹¹ 87%	Epstein-Barr Virus (EBV), innerhalb des BERF4 (= BamHI E fragment rightward reading frame 4); (Baer et al., 1984)
U 379-400	5.0x10 ⁻¹⁰ 91%	EBV, innerhalb des 12-fach repetitiven BamHI W Fragments, im BWRFL
L 163-197	9.7x10 ⁻¹¹ 83%	EBV, innerhalb des BVRFL, mit nicht perfekter (CCCCTG) ₉ Repetition
L 333-359	5.5x10 ⁻¹² 85%	EBV, 3' zum BARFL
	5.7x10 ⁻¹⁰	

Dazu kommen die in Tab. 7 aufgeführten Homologien zu pCaOri5.01 und pCaOri7.01.
Bei U 1-39 handelt es sich ja um den 3'-Teil eines Hamster 4.5S RNA Gens.

Tabelle 10: Homologien von p9 mit bekannten Sequenzen (zu Kapitel 4.3.7.8.)
Die Legende zu Tabelle 9 gilt entsprechend.

4.3.7.8.

Datenbank Sequenzen mit Homologie zu TC7p9 oder p9

Durch eine Computer gestützte Suche in der MicroGenie™ Datenbank (Beckman Instruments, Inc.)* fand ich zahlreiche TC7p9 und p9 homologe Sequenzen. Die signifikantesten Homologien sind in den Tabellen 9 und 10 aufgeführt.

Das Auffallende an diesen Homologien ist bei TC7p9:

1. TC7p9 zeigt zu einer Sequenz 120 bp 5' zum kodierenden Bereich der *C γ* Immunglobulin Genregionen eine Homologie, die man statistisch nur 1x in einer 1.5×10^{11} Nukleotide langen DNA erwarten würde.
2. TC7p9 zeigt Homologien zu den Herpesviren HSV-1, HSV-2 und EBV.
3. Der TC7p9 Bereich 1-250 ist zu keiner bisher bekannten Sequenz in größerem Maß homolog.
4. Der TC7p9 Bereich 250-350 zeigt Homologien zu Sequenzen, die 5' zum cytoplasmatischen β -Aktin Strukturgen des Huhns liegen sowie zu Sequenzen der 4x 500bp terminalen Repetition und des BdRF1 (Pos. 148707-149741; BamHI d fragment rightward reading frame 1) des Epstein-Barr Viruses.
5. Der TC7p9 Bereich von 472-508 mit dem GC₃-Cluster zeigt Homologie zu pCaOri7.01, zu Sequenzen von repetitiven Regionen der Genome von HSV-1 und EBV sowie zu Sequenzen der 5'-untranskribierten Region eines *X.borialis* rRNA Gens und des Gens für die cytoplasmatische Thymidinkinase des Huhns.
6. Der TC7p9 Bereich 600-950 zeigt Homologien zu den Introns einiger Gene.

Bei den Homologien von p9 fällt auf:

1. Die Homologien mit der geringsten Wahrscheinlichkeit eines

*Quelle der Nukleinsäure Sequenzen: GenBank, c/o Bolt, Beranek & Newman, 10 Moulton Street, Cambridge, MA 02238

- zufälligen Auftretens bestehen zu pCaOri7.01 und zu einem *X.laevis* rRNA Gen (zum ETS und, damit verbunden, zum ITS).
2. Die Homologie zu den Herpesviren HSV-2 und EBV ist noch größer als bei TC7p9.
 3. Der p9 Bereich 177-254 (mit dem GC₃-Cluster) zeigt, außer zum *X.laevis* rRNA Gen, noch Homologie zu Sequenzen der 5'-untranskribierten Bereiche des menschlichen Dihydrofolatreduktase Gens und des HSV-2 IE3 Gens.

Die Homologien von TC7p9 und p9 mit anderen, bekannten Sequenzen könnten unter Umständen auf mögliche Funktionen dieser Genom Abschnitte hinweisen. In einigen Fällen mag es lohnend sein, diesen Hinweisen nachzugehen.

4.3.7.9.

Homologie von TC7p9 mit Sequenzen der Immunglobulin Gene

Der TC7p9 Bereich 888 bzw. 906 bis 939 zeigt ausgeprägte Homologie zu einer Intronsequenz, die ca. 80 bp vor dem Exon 1 der *Gamma* Immunglobulin Gensegmente endet. Dies und die beiden Homologien zu Ig κ veranlaßten mich zu der Suche nach Sequenzmotiven (siehe Anhang 7.2.), wie sie für die Enhancerregionen und für das DNA Rearrangement (joining, switching) der Immunglobulin Gene charakteristisch sind.

Konservierte Heptamere und Nonamere sind die Signalsequenzen für das Verknüpfen der Segmente der Immunglobulin Gene (Übersichtsartikel: Tonegawa, 1983) und auch der T-Zell Rezeptor Gene (Chien et al., 1984).

Bei Pos. 522 der TC7p9 Sequenz findet sich das Heptamere CACTGTG und bei Pos. 577 das Nonamere ACAAAAACA. Das Heptamere stimmt völlig, das Nonamere bis auf ein Nukleotid mit den Konsensussequenzen dieser Signalsequenzen überein. Ein Vergleich mit den Segmenten der Immunglobulin Gene zeigt eine gewisse

Übereinstimmung der TC7p9 Region 462-585 mit einem J-Segment:

585	TGTTTTTGT	N ₄₉	CACTGTGCAGCGG	N ₃₈	CCAGCCGTGGTGACAGG	462
	Nonamer		Heptamer	kodierender	Donor	Spleißstelle
				Bereich		

Zwar tritt eine Konstellation: 8/9 identisches Nonamer, bis zu 100 bp Zwischenraum und vollständig identisches Heptamer nicht allzu häufig auf [$P(7) \times P(8/1) \times 100 = 6.6 \times 10^{-7}$], doch gegen einen Zusammenhang zwischen TC7p9 und einem Immunglobulin- oder T-Zell Rezeptor Gensegment spricht, daß in diesem spekulativen Fall ausgerechnet der GC₃-Cluster größtenteils für das J-Peptid kodieren würde und, daß die Sequenz zwischen den Oligomeren bei den Immunglobulin Genen und beim Gen für die β -Kette des T-Zell Rezeptors nur 12 oder 23 ± 1 bp lang ist.

Bei TC7p9 Pos. 1065 findet sich das Heptamere CACAGTC, bei Pos. 1031 das Nonamere CCAATAACC, der Zwischenraum ist 25 bp lang. Das Heptamere ist, bei einer Abweichung von der Konsensussequenz, identisch mit dem J_{H1}-Segment der Maus (Sakano et al., 1980). Das Nonamere hat zwei Abweichungen von der Konsensussequenz. Die beiden Oligomere zeigen jedoch eine - im Vergleich zu den Signalsequenzen für das Verknüpfen der Immunglobulin Gensegmente - entgegengesetzte Orientierung.

Bei TC7p9 Pos. 1076 liegt eine Nukleotidfolge, die perfekt mit der Konsensussequenz SV40 ähnlicher Enhancerelemente (Anhang 7.2.2.) übereinstimmt. Auch die TC7p9 Sequenz 321-333 zeigt dazu Homologie (siehe auch 4.3.7.4.).

1076	CTTACCACA	1084	TC7p9
	CTTACCACA		MSV Enhancer Stammsequenz
	CAAAACCACA		Ig _H Enhancer Stammsequenz 1
167	ACTTTCACACC	178	SV40
321	ACTCCCCACACC	333	TC7p9

Bei TC7p9 Pos. 936 findet sich, wie schon gezeigt (4.3.7.6.) eine Homologie mit einem weiteren Transkriptionskontrollelement.

```

934 cccTGCATATctccctca 951 TC7p9
133 cttTGCATACTttctgcct 150 SV40
166 attTGCATTTctaaaata 183 IgH Enhancer
-66 attTGCATA    ccctca -52 IgK Promotor

```

4.3.7.10.

Homologie zwischen p9 und einem *X.laevis* rRNA Gen

p9 zeigt große Homologie zu den Bereichen eines *X.laevis* rRNA Gens, die die Sequenz der 18S RNA flankieren: zum externen- und zum 1. internen transkribierten Zwischenraum (ETS und ITS1) der rRNA Transkriptionseinheit (40S rRNA Vorläufer).

Diese beiden Bereiche haben in hohem Maß die Fähigkeit, Sekundärstrukturen auszubilden, sind G+C reich und spielen vermutlich in der Prozessierung der 18S rRNA eine Rolle. In einer Modellvorstellung würden die zentralen Bereiche von ETS und ITS1 Sekundärstrukturen ausbilden; die Stellen, an denen geschnitten wird, blieben einzelsträngig.

Besonders auffallend ist, daß sich die Homologien zwischen p9 und ITS1 durchgehend anordnen lassen:

p9	181-241	290-328	359-414	404-465
ITS1	2992-3052	3120-3153	3230-3180	3228-3289

Die ITS1 RNA hat selbstkomplementäre Sequenzen, die 3 große Haarnadelschleifen ausbilden können. p9 dagegen besitzt keine selbstkomplementären Sequenzen. Homologie von p9 zu ITS1 besteht entweder nur zu einem Strang der Haarnadelschleifen,

ist durch nicht komplementäre Basen eingeschränkt oder zeigt abschnittsweise (359-414 zu 3230-3180) die entgegengesetzte Orientierung (Abb. 27).

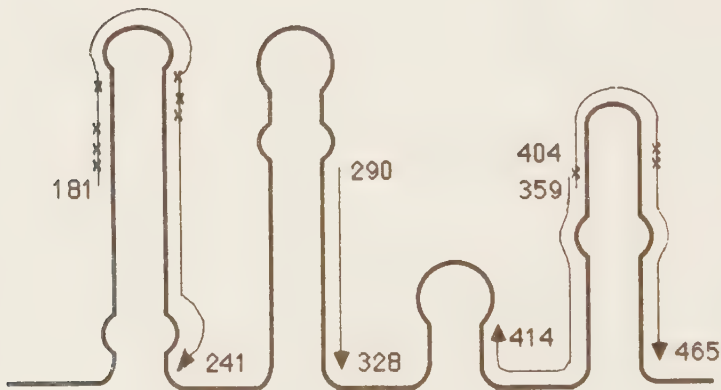


Abbildung 27: Homologie zwischen p9 und der ITS1 Region eines *X. laevis* rRNA Gens

Die Sekundärstruktur des ITS1 einer *X. laevis* 40S rRNA (Pos. 367-681; Hall & Maden, 1980) ist durch die dicke Linie dargestellt. Dünne Linien mit Pfeilen bezeichnen homologe p9 Sequenzen und deren Polarität. An den Stellen wo p9 Homologie zu beiden Strängen einer Haarnadelschleife hat, zeigen Kreuze die unterschiedlichen Basen an, die in einer Haarnadelschleife zu nicht komplementären Paaren führen würden.

Es scheint demnach eine Systematik in der Beziehung der beiden knapp 300 bp langen Sequenzen zu liegen: Die Selbstkomplementarität, die bei ITS1 auf der Ebene der RNA möglicherweise eine entscheidende Rolle spielt, tritt bei p9 nicht auf.

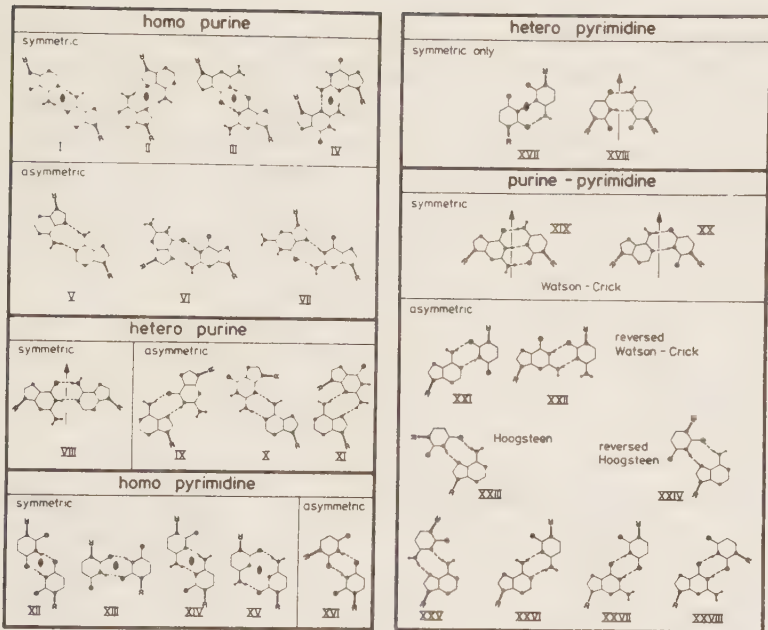


Abbildung 28: Die 28 möglichen Basenpaare aus A, G, U(T) und C mit mindestens 2 zyklischen Wasserstoffbrückenbindungen (aus Saenger, 1983)

Bei den Anordnungsmöglichkeiten VI, VIII, IX, XVI, XVIII, XIX, XX, XXIII und XXVIII befinden sich die glykosidischen Bindungen (dicke Linien zu R) auf der gleichen Seite des Basenpaars, was Voraussetzung für die Bildung einer Doppelhelix mit antiparallelen Strängen ist. Die Paare VI, IX, XIX, XX und XXVIII zeigen fast gleiche Abstände der C1'-Atome (R). (Die Symbole ● und ↑ bezeichnen 2-zählige Drehachsen senkrecht zur bzw. in der Papierebene.)

Davon gibt es 9 Anordnungsmöglichkeiten, wo sich beide glykosidischen Bindungen auf der gleichen Seite des Basenpaares befinden. Diese Voraussetzung für die Bildung einer Doppelhelix mit antiparallelen Strängen ist auch im Fall einer Haarnadelschleife zu fordern.

Folgende 5 Nukleotidpaare (VI, IX, XIX, XX und XXVIII; Abb. 28) zeigen dazu noch fast gleiche Abstände der beiden Cl'-Atome (ca. 107nm) und wären, bei (der gewöhnlichen) anti/anti Konfiguration der Paare G-C, A-T und G-T sowie syn/anti Konfiguration der Paare G-G und A-G, mit einer Doppelhelix DNA Struktur vereinbar:

G-C	anti/anti	Watson-Crick Basenpaarung
A-T	anti/anti	(= A-U) Watson-Crick Basenpaarung
G-T	anti/anti	(= G-U) Kodon/Antikodon Erkennung
G-G	syn/anti	Hoogsteen Basenpaarung
A-G	syn/anti	Hoogsteen Basenpaarung

Eine Hoogsteen Basenpaarung beinhaltet eine Wasserstoffbrückenbindung zum N7 eines Purins und wurde in Kristallen des 1:1 Komplexes aus 1-Methylthymine und 9-Methyladenine erstmals nachgewiesen (Hoogsteen 1959, Hoogsteen 1963). Daß Hoogsteen Basenpaarung auch in Nukleinsäuren vorkommt, zeigt das Beispiel poly(dA)·poly(dT), wo sich, wenn man die relative Feuchtigkeit vermindert, die B-DNA ähnliche Struktur nicht in eine A-DNA Struktur umwandelt, sondern sich eine Tripelhelix poly(dA)·2poly(dT) bildet. Diese besteht aus einer poly(dA)·poly(dT) Doppelhelix mit A-DNA Struktur, in deren großer Furche ein weiterer poly(dT) Strang über Hoogsteen Basenpaarung parallel zum poly(dA) Strang angeordnet ist (Arnott & Selsing, 1974). Die RNA Systeme poly(A)·poly(U) und poly(I)·poly(A) sind dazu analog.

Die Bezeichnungen anti oder syn beziehen sich auf die Stellung des Zuckers und der Base zueinander. Dabei betrachtet man den Torsionswinkel χ (Chi) zwischen den Bindungen der Atome

04'- Cl' - N1 - C2 bei Pyrimidinen, bzw.

04'- Cl' - N9 - C4 bei Purinen (Abb. 29).

Nach der Nomenklatur von Klyne und Prelog (IUPAC, 1971), steht

syn allgemein für einen Wert des Torsionswinkels Θ (Theta) von $0^\circ \pm 90^\circ$ (synperiplanal - synclinal), anti für $180^\circ \pm 90^\circ$ (antiperiplanar - anticlinal).

In Lösung stehen die Konformationen syn und anti im Gleichgewicht. Bei Purin Nukleosiden sind beide Konformere ungefähr gleich häufig, in Pyrimidin Nukleosiden überwiegt anti. Dies hat unter anderem mit der verschiedenen großen sterischen

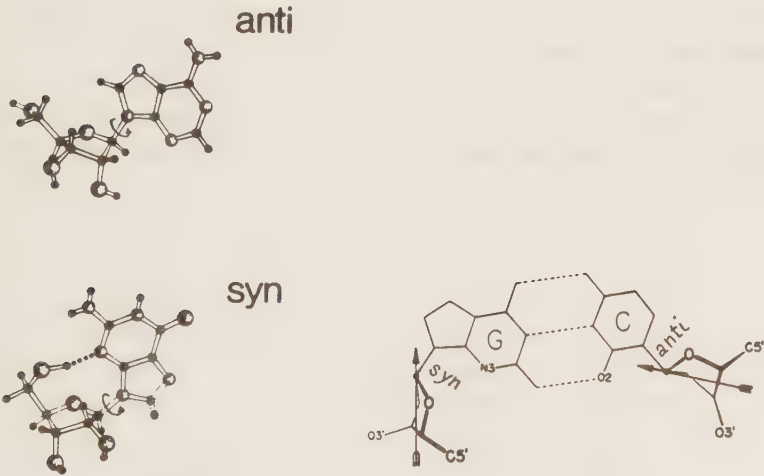


Abbildung 29: syn/anti Konfiguration zwischen Zucker und Base (aus Saenger, 1983)

Die beiden hauptsächlichen Orientierungen von Zucker und Base zueinander werden syn und anti genannt und sind durch den Torsionswinkel - χ (Chi): $O4'-C1'-N1-C2$ (bei Pyrimidinen) bzw. $O4'-C1'-N9-C4$ (bei Purinen) - definiert (Pfeile). Adenosin ist in der anti-, Guanosin ist in der syn Konformation dargestellt. In der Z-DNA kommt die syn Orientierung durch die im Vergleich zur A- oder B-DNA veränderte Lage des Zuckers zustande (unten rechts). Bei einer Hoogsteen Basenpaarung $A_{syn}-G_{anti}$ oder $G_{syn}-G_{anti}$ dagegen könnten die Zucker eine Stellung wie in A- oder B-Helices einnehmen bzw. beibehalten.

Hinderung zu tun. (6-Ring gegenüber 5-Ring, O2-Sauerstoffatom gegenüber N3-Stickstoffatom.) Das syn:anti Verhältnis ist durch Substituenten beeinflusst.

Guanosin Nukleotide stellen bezüglich der syn/anti Konformation eine Ausnahme dar. Bei Guanylsäure (GMP) überwiegt in der Lösung die syn Konformation, und dGMP kristallisiert als Monomeres oder innerhalb einer Z-DNA Struktur in der syn Form. Verantwortlich für diese Bevorzugung der syn Konformation sind van der Waals und elektrostatische Wechselwirkungen zwischen der Aminogruppe an C2 und der 5'-Phosphatgruppe.

In der Z-DNA Struktur findet man ausschließlich Watson-Crick Basenpaarung, die syn Stellung von Chi kommt durch die im Vergleich zur A- oder B-DNA veränderte Lage des Zuckers zustande (Abb. 29). Bei einer Hoogsteen Basenpaarung $A_{syn}-G_{anti}$ oder $G_{syn}-G_{anti}$ dagegen könnten die Zucker eine Stellung wie in A- oder B-Helices einnehmen bzw. beibehalten.

5.

DISKUSSION

5.1

Potentielle regulatorische
Elemente des c-myc Gens

Nach der Isolierung des c-myc Gens der Affenart *Cercopithecus aethiops* aus dem Genom der TC7 Zelllinie stehen nunmehr die c-myc Gene vier verschiedener Vertebraten zur Verfügung.

Die Ermittlung der Nukleotidsequenz TC7myc1 (Exon 1 mit flankierenden Regionen) ermöglichte den Vergleich dieses für die Regulation wichtigen Genabschnitts mit den c-myc Genen von Mensch, Maus und Huhn.

TC7myc1 ist zu den entsprechenden Sequenzen des Menschen und der Maus zu 94% bzw. 69% homolog. Die betreffende Sequenz des Huhns ist dazu nur sehr entfernt verwandt. Zwischen diesen 5'-Bereichen der verschiedenen c-myc Gene besteht somit weit weniger Homologie als zwischen den kodierenden Exons.

Exon 1 ist etwas besser konserviert als die 5'-untranskribierte Region und Intron 1. Erhalten geblieben ist auch die Genstruktur mit der ungewöhnlich langen 5'-Führungssequenz, die sich auch beim N-myc Gen findet. Der Grund für diese Konservierung sind wahrscheinlich Funktionen in der Regulation der Genexpression.

Eine Funktion mancher ungewöhnlich langer 5'-untranslatierter Regionen (Kozak, 1984) ist die Regulation der Expression über differentielle oder selektive Translation.

So verstärkt die dreiteilige Führungssequenz der vom hauptsächlichen späten Adenoviruspromotor (Ad MLP) ausgehenden Transkripte die Translation nur im späten Stadium der Infektion (Logan & Shenk, 1984). In der Hefe gibt es das gcn4 Gen, dessen Expression auf translationeller Ebene reversibel blockiert wer-

den kann (Thireos et al., 1984; Hinnebusch, 1984). Schließlich die Hitzeschockgene, deren Transkripte in hitzebehandelten Zellen bevorzugt translatiert werden (Scott & Pardue, 1981). Dafür ist, wie bei hsp70 (McGarry & Lindquist, 1985) und hsp22 in *Drosophila* nachgewiesen, die konservierte Sequenz am Beginn der mRNA (Hultmark et al., 1986) verantwortlich. Bei hsp22 wurde zudem gezeigt, daß dieser Abschnitt auch noch einen aktivierenden Einfluß auf die Transkription hat.

c-myc Transkripte in transformierten Zellen haben häufig andere cap-Stellen und andere Führungssequenzen als im Normalfall; dabei fehlt Exon 1 oft teilweise oder ganz. Für v-myc gilt entsprechendes.

Deshalb wurde allgemein eine Regulation von c-myc durch differentielle Translation in Erwägung gezogen. Sekundärstrukturen der mRNA innerhalb Exon 1 (Battey et al., 1983) oder zwischen Exon 1 und Exon 2 (Saito et al., 1983), wurden als Elemente, die die effektive Translation behindern könnten (siehe dazu etwa: Pelletier & Sonenberg, 1985), ins Spiel gebracht.

Untersuchungen *in vivo* (Nilsen & Maroney, 1984) bzw. mit transfizierten Konstrukten (Butnick et al., 1985) haben jedoch eine vergleichbar effektive Translation der verschiedenen c-myc mRNAs ergeben. *In vitro* Translation *in vitro* transkribierter c-myc Konstrukte, ausgeführt von 2 anderen Gruppen (Persson et al., 1984; Darveau et al., 1985), ergaben einander widersprechende Ergebnisse, obwohl die Versuchsansätze nahezu gleich waren.

Durch vielfältige Sequenzvergleichsanalysen fand ich in der 5'-Region des c-myc Gens einige DNA Abschnitte die möglicherweise für die Expression des Gens von Bedeutung sind:

- Das Sequenzmotiv GTTTACAT, identisch mit dem Enhancer Element des *X. laevis* snRNA U1a Gens und homolog zu zahlreichen anderen, gut charakterisierten Transkriptionskontrollelementen (Abb. 9).
- Der pyrimidinreiche, repetitive TCCCCA Bereich mit der DNase

hypersensitiver Stelle III₁ (Siebenlist et al., 1984) und mit Homologie zum SV40 Enhancer und zum Promotor des Hühner β -Aktin Gens (Abb. 10).

- Die Spl Konsensussequenz (Abb. 11).

- Der Bereich um die DNase hypersensitive Stelle III₂, mit inverser Repetition, Homologie mit dem hsp83 Hitzeschockelement, Homologie mit der Sequenz 169-147 der SV40 DNA und potentieller T-Antigen Bindungsstelle (Abb. 12).

- Die zwischen Mensch, Affe und Maus vollständig konservierten 26 Nukleotide nach der cap-Stelle 2 (4.2.6.5.).

Es gibt eine Verbindung zwischen den myc und den hsp Genen: c-myc kann, wie das Onkogen AdEla (Nevins, 1982), das hsp70 Gen induzieren (Kingston et al., 1984). Und, hsp wie myc Gene können durch Stimulation der Zelle mit Serum oder Wachstumsfaktoren (Wu & Moromoto, 1985; Kelly et al., 1983; Campisi et al., 1984) und auch entsprechend dem Entwicklungsstadium (Zimmerman et al., 1983; Zimmerman et al., 1986) induziert werden. Das für die Stimulierung durch Serum verantwortliche Element des menschlichen hsp70 Gens ist als eine Purinnukleotidfolge, 58 bp 5' zur cap-Stelle, kartiert worden (Wu et al., 1986; Anhang 7.2.6). In vergleichbaren Positionen finden sich auch im c-myc Gen Purinnukleotidfolgen (Abb. 6).

Mechanismus, Faktoren und Elemente der c-myc Expression sind bisher erst ansatzweise bekannt. Ein gegenwärtig favorisiertes Modell ist die konstitutive Transkription des c-myc Gens und die Regulation über den Abbau der mRNA (Dani et al., 1985; Piechaczik et al., 1985; Blanchard et al., 1985; Dony et al., 1985).

Als Elemente eines konstitutiven Promotors kommen das Sequenzmotiv GTTACAT, der TCCCCA Bereich und die Spl Konsensussequenz in Betracht. Eine Analyse entsprechender Mutanten könnte dies klären. Zur Identifizierung von Faktoren, die spezifisch an

c-myc Promotorelemente binden, bieten sich Gel Retentionstests (Schneider et al., 1986) mit chemisch synthetisierten Oligonukleotiden der potentiellen Transkriptionskontrollelemente an.

Wie an primären Milzzellen und 3T3 Zellen der Maus getestet (Kelly et al., 1983), führt die Inhibition der Protein Biosynthese durch Cycloheximid zu einer erhöhten Menge an beiden c-myc mRNAs. Dies ist nach den Versuchen von Dani et al. (1984) an einigen menschlichen Zelllinien auf die Stabilisierung der extrem labilen c-myc mRNA zurückzuführen und würde bedeuten, daß ein wiederum instabiles Protein existiert, welches, direkt oder indirekt, spezifisch den Abbau der c-myc mRNA fördert. Eine Zielstruktur für den postulierten Nukleasfaktor liegt eventuell auf Exon 1, denn fehlt dieses ganz, oder zum größten Teil, ist die c-myc mRNA um den Faktor 3-5 stabiler (Piechaczik et al., 1985; Butnick et al., 1985).

Die inverse Repetition zu Beginn der mRNA mit der cap-Stelle 2 käme beispielsweise für diese Zielstruktur in Frage.

Sicher sind bereits Experimente hinsichtlich dieses Nukleasfaktors im Gange, der ein Schlüssel für das Verständnis der c-myc Regulation wäre.

5.2.

p9 und TC7p9, zwei weitere zelluläre Sequenzen mit Homologie zum regulatorischen Bereich der SV40 DNA

Im Genom der TC7 Zelllinie gibt es für die für Maus, Ratte und Hamster spezifische 4.5S RNA wie für das gesamte p9 Hamster DNA Insert kein Äquivalent.

Da aber der Bereich von Pos. 40-529 des p9 Inserts eine ungewöhnliche Basenzusammensetzung hat und ein Transkriptionskon-

trollelement enthält, habe ich, um entfernt verwandte, aber möglicherweise funktionell ähnliche Regionen im Genom des Affen zu finden, die TC7 Genbank mit p9 durchgemustert und 4 verschiedene rekombinante Phagen mit überlappenden Inserts isoliert, die einen 27kb Genomabschnitt darstellen.

Gleichartige Versuchsansätze mit der regulatorischen Region der SV40 DNA als Probe für das Durchmustern einer Genbank haben zur Isolierung von SV40 ähnlichen Sequenzen aus dem Genom des Affen (pCaOri Klone; McCutchan & Singer, 1981) und des Menschen (Conrad & Botchan, 1982) geführt. Dadurch wurden, besonders im Falle von pCaOri7.01, DNA Abschnitte mit äußerst interessanten Eigenschaften (Saffer & Singer, 1984; Gidoni et al., 1984; Dynan et al., 1985) ans Licht gebracht.

Der zu p9 (HTC) homologe Bereich des 27kb Genomabschnitts der TC7 DNA wurde schrittweise auf ein 336bp SacI Fragment eingegrenzt. Dabei zeigte sich ein Hybridisierungsphänomen: nur einer der beiden komplementären Stränge des SacI Fragments hybridisierte mit p9.

Um die Homologie zwischen p9 und der TC7 DNA zu bestimmen, um herauszufinden, was die Funktion des isolierten Genomabschnitts sein könnte und nicht zuletzt, um dem Hybridisierungsphänomen nachzugehen, habe ich ein 1.13kb PstI Fragment (TC7p9), in dessen Mitte sich das 336bp SacI Fragment befindet, sequenziert.

Grund für die Hybridisierung zwischen p9 und TC7p9 sind die 26 bzw. 24 bp langen GC₃-Cluster mit 92% Homologie zueinander. Ein zu p9 89% und zu TC7p9 84% homologer 28bp GC₃-Cluster findet sich auch in pCaOri7.01.

Der Vergleich der Basenzusammensetzung und zahlreiche Sequenzvergleichsanalysen zeigten weitere Gemeinsamkeiten zwischen p9, TC7p9, den pCaOri Klonen und der regulatorischen Region der SV40 DNA. Deswegen kann man p9 und TC7p9 als weitere SV40 ähnliche Sequenzen klassifizieren.

Die GC₃-Cluster zeigen Homologie zu GC-reichen Sequenzen der 5'-untranskribierten Regionen folgender Gene:

Dihydrofolatreduktase (Mensch), β -Aktin und Thymidinkinase (Huhn) und IE3 (HSV-1 und HSV-2).

Diese Regionen enthalten allesamt potentielle Bindungsstellen für den Spl Transkriptionsfaktor. Spl Bindung an Promotorsequenzen des HSV-1 IE3 Gens und des Dihydrofolatreduktase Gens der Maus ist vor kurzem nachgewiesen worden (Jones & Tjian, 1985; Dynan et al., 1986).

pCaOri7.01 enthält 2 Regionen, die Spl binden (Gidoni et al., 1984); pCaOri5.01 und pCaOri9.32 besitzen potentielle Spl Bindungsstellen. p9 und TC7p9 enthalten einige Male die kanonische Spl Stammsequenz (GGGCGG) aber nicht die ganze Konsensussequenz der Spl Bindungsstellen (Übersichtsartikel: Kadonaga et al., 1986). Allerdings ist die p9 Sequenz 432-441 identisch mit der HTLV-III Spl Box III (Jones et al., 1986).

Eine weitere Besonderheit der GC₃-Cluster ist ihre Verknüpfung mit Sequenzen, die aus alternierenden Purin- und Pyrimidinnukleotiden bestehen.

Sequenzen mit alternierenden P/Q sind in den Genomen offensichtlich erheblich unterrepräsentiert (Anhang 7.3.). Ausnahmen sind dabei anscheinend virale Promotorregionen. Dies könnte funktionelle Bedeutung haben, worauf auch das Vorkommen des Sequenzmotivs TGCAQP in Transkriptionskontrollelementen hindeutet (Abb. 25).

Auch das Hybridisierungsphänomen ist eine Parallele zwischen p9/TC7p9 einerseits und pCaOri9.32/SV40 andererseits.

Das Hybridisierungsphänomen ist wahrscheinlich ein Effekt der besonderen Basenzusammensetzung, die es nur einem der beiden komplementären Stränge ermöglicht, eine Sekundärstruktur mit zum Teil ungewöhnlichen Basenpaarungen zu bilden.

Künftige Experimente (z.B. Tests auf S1 Nuklease Sensitivität)

müssen zeigen, ob diese Struktur nicht nur unter den Bedingungen der Southern Blot Hybridisierung, sondern auch *in vivo* vorkommt. Auch müssen die betreffenden DNA Abschnitte im Hinblick auf eine mögliche funktionelle Bedeutung untersucht werden, was vielversprechend erscheint, denn die GC₃-Cluster - mit ihren Besonderheiten - sind am Hybridisierungsphänomen beteiligt.

So ist auch die Funktion des gesamten 27kb DNA Abschnitts mit der TC7p9 Sequenz noch offen, zumal es mir bisher nicht gelungen ist, im Bereich des Klons pHK6.1 Transkription nachzuweisen. Hier bedarf es zur Abklärung weiterer Experimente. Durch den Computer unterstützten Vergleich (Queen & Korn, 1984) mit einer Sequenzdatenbank (Beckmann MicroGenie™) wurden die signifikantesten Homologien von TC7p9 und p9 mit bekannten Nukleotidsequenzen dokumentiert. Der Computer unterstützten Sequenzvergleichsanalyse, wie sie innerhalb dieser Arbeit exemplarisch durchgeführt wurde, kommt, bei der rasch zunehmenden Fülle von Sequenzdaten, eine immer größere Bedeutung zu. TC7p9 ist offensichtlich eine bisher unbekannte Region im Genom der Primaten. Die gefundenen Homologien sind als Hinweise auf die Funktion und die Herkunft des 27kb Abschnitts der TC7 DNA anzusehen. Es gilt nunmehr, aus diesen Hinweisen die richtigen Schlüsse zu ziehen.

6.

ZUSAMMENFASSUNG

Von der TC7 Zelllinie der Affenart *Cercopithecus aethiops* wurde eine Genbank im Phagenvektor Lambda 1059 angelegt. Das aus der Genbank isolierte c-myc Gen wurde durch Restriktionsanalyse charakterisiert und steht nunmehr für funktionelle Experimente zur Verfügung. Der 5'-untranskribierte Bereich, das Exon 1 und ein Teil des 1. Introns wurden sequenziert (TC7myc1) und durch Computer unterstützte Sequenzvergleichsanalyse untersucht. Auf diese Weise wurden potentielle regulatorische DNA Sequenzelemente gefunden.

Daneben wurde die TC7 Genbank mit p9 (einer Integrationsstelle der Ad12 DNA ins Genom einer Hamster Tumorzelllinie, die ein Transkriptionskontrollelement enthält) als Probe durchgemustert, um entfernt verwandte, aber möglicherweise funktionell ähnliche Regionen im Genom des Affen zu finden. Dies führte zur Isolierung eines 27kb Genomabschnitts. Der zu p9 homologe Bereich wurde auf ein 336bp SacI Fragment eingegrenzt. Dabei trat ein Hybridisierungsphänomen auf: nur einer der beiden komplementären Stränge hybridisierte mit p9. Die Sequenzierung eines 1.13kb PstI Fragments (TC7p9) zeigte, daß p9 und TC7p9 nahezu identische GC₃-Cluster, Sequenzen mit dem repetitiven Tetranukleotid GCCC, enthalten. Durch zahlreiche Sequenzvergleichsanalysen wurden so viele Gemeinsamkeiten zwischen p9, TC7p9, den pCaOri Klonen (McCutchan & Singer, 1981) und der regulatorischen Region der SV40 DNA offenbar, daß es gerechtfertigt ist, p9 und TC7p9 als weitere SV40 ähnliche Sequenzen zu klassifizieren. Nicht zuletzt zeigt auch pCaOri9.32 ein Hybridisierungsphänomen. Diese Parallele ermöglichte es mir, eine Hypothese zur Erklärung des Hybridisierungsphänomens aufzustellen. Auch wurden die signifikantesten Homologien von p9 und TC7p9 mit Datenbanksequenzen und mit funktionellen DNA Sequenzelementen dokumentiert.

7.

ANHANG

7.1.

Zum Vergleich herangezogene
Sequenzen

7.1.1.

c-myc Gen des Huhns

Exon 1 und flankierende Sequenzen (Shih et al., 1984)

CHICKEN CELLULAR MYC PROTO-ONCOGENE, COMPLETE CDS, 801-1800

```

GGAGCGCCCT TTTCATCCCC GCTCGTATTT TTTTTTTTTT TTTACTATGT 850
TTACTTCCGA CCTCTCCTTG TAGTAGGAAA AAAAACCAAC CGCTGCTCCG 900
CATCGCCTCT CCCCggcccc TCTCCCTCCC TCCCTCCCTC CCGCCCCGCC 950
AGCTCCGGCT CGCAGTACTC GGGGGGGGGG ACggGAGCCCC TCGGCCGCC 1000
CCTCGCGGCG CGCCCTCCCC GCTCACGGAG CCCGCGCGGA GCCGGGGGGCG 1050
AGCGGGAGGG AGATGAAGCG GCGACGCGCA CCGCGAGAGC GCGCACTCGC 1100
GGGGCCCCGC CGTGCCGCTC GTGCTCCCGC CCCCgCTGCA TCTCCCGCCC 1150
GCCCTCGCC GCGGCTTTAA AGACAGCAA GCAACTTAAC TTCTATTGTA 1200
CCGACGGAG CGCGCCGCGC CGCCTTGGAC CGTACAATCT GCCGCACGCC 1250
GGGAAGGCGA GCCCTCTCCG CTGTATTTTT TTTCTCATCG TGGTGGGAGG 1300
AAGCGATCTA CGTTCTCCCG GCGTTTTGGC TCCCTTTTTT CCCCCGTCTT 1350
CTCCGGCGGT TCTTTTTATA TTTTATCTC CAATTTCTGT ATTTGTTGT 1400
TCCCGCACC GCCCGCAATT ATTGCCTCGC TCGTCTCGC AGCCGGCGCG 1450
TTGGAGGAGC CGGTGAGTGG CGGGGCTCCG CTTTACCCAT CACTCGATCC 1500
GCCGCGGTGCG GGCTGACAGC GCGGGGCCCG GGGACGCGG GTGCCTCCCG 1550
GGGACGCGG CGCTGCTCCG GGCACGGCGC GCCCCATCCT TCTCCGTGCT 1600
CTCGGCTTGG ATATATAATT CTATTTTTTG GAGGGGGGGG GGGGGGTGGG 1650

```

GGAGGGCAAG AAGCATTTCG TTCTCCAGCG TACGCACGGC AGGTTATGCT 1700
 TATTGCACAT ATATACGTAT ATATATGTGC GTGTGTGGAT ATATATGTAT 1750
 ATATGATAAA TTTGGCAAAG TTTGCCCAGC TCCGTGCACG GCCAGGTGGG 1800

7.1.2.

N-myc, Exon 1 (Kohl et al., 1986)

TTGTCTGCGC CACAGCCCCC TTCTCTCCCC GCCCCCGGG TGTGTCAGAT 50
 TTTTCAGTTA ATAATATCCC CCGAGCTTCA AAGCGCAGGC TGTGACAGTC 100
 ATCTGTCTGG ACGCGCTGGG TGGATGCGGG GGGCTCCTGG GAACTGTGTT 150
 GGAGCCGAGC AAGCGCTAGC AGGCGCAAGC GCGCACAGAC TGTAGCCATC 200
 CGCGGACACC CCGCCCCCCC CGGCCCCACC GGAGACACCC GCGCAGAATC 250
 GCCTCCGGAT CCCCTGCAGT CGGCGGGAGG TAAGGAGCAG GGCTTGCAAA 300
 CCGCCCCGGC CCCAGGGAAG CGACGAGCGC CGGGGCAAGG CAAGCCCTGG 350
 ACGGGATTGC GACGTGCGCA CCGGGCGCCC TAATATGCCC GGGGGACTGT 400
 TTCTGCTTCC GAAACAAAAC CATCTCTGGG TTTTCCCAGA AAAGCCAGTT 450
 CCAGCCCCGA AGGCATCCTG GCTAGAGGAG ACCCGCCCTA ATCCTTTTGC 500
 AGCCCTTACC GGGGGGAGTA ATGGCTTCTG CGAAAAGAAA TTCCCTCGGC 550
 TCTAGAAGAT CTGTCTGTGT TTGAGCTGTC GGAGAGCCGG TCGTCCCCCA 600
 CCCTAGGCTG GGGTTCTTCT CCAAAGGGTG CCC 633

7.1.3.

Regulatorische Region der SV40 DNA

SV40, 5164-334 (Buchman et al., 1981)

CTTTGCAAAG CTTTTTGCAA AAGCCTAGGC CTCCAAAAAA GCCTCCTCAC 5213
 ◀----- früh ----- T-Ag I -----

-----▶ ◀----- TATA Box
 TACTTCTGGA ATAGCTCAGA GGCCGAGGCG GCCTCGGCCT CTGCATAAAT 20
 ----- T-Ag II -----

Spl Box I Spl Box II Spl Box III
 AAAAAAAATT AGTCAGCCAT GGGGCGGAGA ATGGGCGGAA CTGGGCGGAG 70
 ----- T-Ag III -----

Spl Box IV Spl Box V Spl Box VI -----
 TTAGGGGCGG GATGGGCGGA GTTAGGGGCG GGACTATGGT TGCTGACTAA 120

Spl I S1 hy'sens.

 TTGAGATGCA TGCTTTGCAT ACTTCTGCCT GCTGGGGAGC CTGGGGACTT 170
 PQPQP QPQ QPQPQ PQ -----▶ -----▶

-----▶----- 72bp Repetition -----
 TCCACACCTG GTTGCTGACT AATTGAGATG CATGCTTTGC ATACTTCTGC 220
 = Enhancer Stammsequenz

-----▶-----
 CTGCTGGGGA GCCTGGGGAC TTTCCACACC CTAAGTGACA CACATTCCAC 270
 PQP QPQPQ

AGCTGGTTCT TTCCGCCTCA GAAGGTACCT AACCAAGTTC CTCTTTTCTAGA 320

spät -----▶
 GGTTATTTCA GGCC 334

7.1.4.

Hamster DNA Anteil von p9 (Stabel & Dörfler, 1982)

```
GATCACGAGT TCGAGGCCAG CCTGGGCTAC ACATTTTTTT TCTCCTCGCC 50
TGTGGCCAGC TGCACTTTTT CGTCTCTCCC GTCTGCGCTC GGCTCCTTTG 100
CTTCGGCCTC TTTTGCTCCG GCCTCGTCTT CAGCCCCGCT GCCCGCTTGC 150
CACAACCCCC CGTTGCCGTT GCCATTGTCC CTGCCTCTGC CCCGTGCAGC 200
CCGCCCCGCC GCCCGCCGCG CCGCACTCCA CCGACGCGGC CTCTCCGCAC 250
GCTTGACAAA GCAGCAACCA GCTCCTCCCG CTTCCTCTGC TCTCCGCGTC 300
CCGCGCTCAA CCGTGGCCCG CCCTCGGGAG ACCCCACCGT GCCCTTCAGC 350
CCTTTCTCCG GCGCCTGCCT CGCCCTGCCT CGCCTCCCCG GCTCCACCAG 400
CAAGCGCTCC CGCCAGACCG ATGCGCCGCA AGCCACGCCT CCTCGTCTC 450
GCAAGCCGCA CGCCGAGGTT CGGCCTCGGT GTCCCCGAGC CTCCCGCGCT 500
CCGGCTCCGA CTTGAAGGAG ACGCCGCCA 529
```

7.1.5.

pCaOri5.01 (Queen et al., 1981)

AFRICAN GREEN MONKEY SV40 HOMOLOGUE REPLICATION ORIGIN

```
CTGCAGGACG CTCCCGCGCA GCCCGGTTTC CCTACGCCGG GGCGGGGCGG 50
AGGGGCGGAG GGGCGGAGGG GCGGACGGGC GGGGTGGGGG AGAGGGCGGA 100
AGGGGCGGGG ACTTCCCTCA GAGGGCGGGG CGGCCGGGCA GCGGCGGCCT 150
CCGACTGGAT GCGGCGGATC GCCGGGGGCC TGAGGGTGA GTCCCGGTCTG 200
GAGGCAGGCC GGGACTCTGG TGGGTCG 227
```

7.1.6.

pCaOri7.01 (Queen et al., 1981)

AFRICAN GREEN MONKEY SV40 HOMOLOGUE REPLICATION ORIGIN; 2

```

TCGACCACAG CCAGAGTCCA TGCATCGGGA GGTTCAC TCG GTTTGCGAAG 50
AACGGGCAGG GCATGCACGG CCTGGGCTCG GCGGGCGGGC GGGCGGGCCG 100
GGGCGCAGTT CCCAGGTTCT CCACTAGAGG TCAGGAGGTG ACCGCTTCGG 150
GGCTGGAAGA CGGGCCCGTC GTGGATTGGC TAGTGCCGGC GGAGGGCGGG 200
GCGGAGAGTG GGGCGGGGCG GAGAGTGGGG CGGGGCGCAG TTCCCCAGTT 250
CGCCACTAGA GGTCAGGAGG TGACCGCTTC GGGGCGGGAA GACTGGCCCCG 300
TCGGGGATTG GCTAGTGCCG GCGGGGGGCG GGGCGGGGGG CGGAGGGCGG 350
GGTGGACGTG GCACCTGGTT GCTGACATCT GGAATGACTT TTTTTTGGCA 400
TCGGATTTC TGTCTTTGTG GGGCTGATGG ACCCGA 436

```

Korrekturen (Saffer & Singer, 1984): nach Pos. 110 ein C, nach
 Pos. 247 ein G einfügen; Pos. 363 = G, Pos. 403 = A

7.1.7.

pCaOri9.32 (McCutchan & Singer, 1981)

AFRICAN GREEN MONKEY HOMOLOGOUS REGION TO SV40 REPLICATION
ORIGIN

```

AGCTTGACTA GGGGCACGGC GGGCGCCTCG CCTGGCGAGG GGCGGGGCGG 50
ACGGGGCGGC CAGCTAGGGG CGGAGCAGGG CGGAGCAGGG CGGGTGCAAG 100
GCGAGTGGTG GGGACCCAGT TGGAAGCGGA GGCCATAGGG TAAGCCCGCC 150
GAGGGGCGGC GGGGCGGCAG GGGAGGCCTG ATGACGACCT GCGCGTGCGG 200
CTCGGCCTGA AGGGGCAAGC AGAGGACTTG GCAGGGCTGG TGAGGGGCGG 250
GGCAGGCGGC GCGGTGGGGG CGGGCCGAGC CCGGAGGTCG GATGAGCGGA 300
CACGGCCCGA CGCGCGAAGC CATGCAAGTA GGTGGCTCCC GACGGCCCCG 350
CTTGAATTTC GATCCCAAAC CGGGTCCNNN CTCCCGGCC 389

```

7.2.

DNA Sequenzelemente

7.2.1.

Adenovirus Enhancer Stammsequenzen

(Hearing & Shenk, 1983; Herbomel et al., 1984)

Ad12 Ela -200	CGGAAGTGAAA	HS ¹
Ad12 Ela -300	CGGATGTGACG	HS
Ad12 Ela -384	AGGAAGTGGGG	
Ad2 Ela -200	AGGAAGTGAAA	
Ad2 Ela -300	GGGAAGTGACA	
Ad2 Ela -380	CGGAAGTGTGA	
Ad2 MLP ² -90	AGGAAGGTGATT	HS
Ad5 Ela -200	AGGAAGTGAAA	HS
Ad5 Ela -300	AGGAAGTGACA	HS
Ad5 Ela -344	CGGATGTGGCA	
Ad5 E4 -150	CGGAAGTGACG	HS
Ad7 Ela -200	AGGAAGTGAAT	HS
Ad7 Ela -300	CGGAAGTAGAC	HS
ASV LTR ³ -100	AGGAAGGCAA	HS
BKV ⁴ 68bp rep. -90	AGGAAAGTGCAT	HS
MMTV ⁵ LTR -160	AGGATGTGAGA	HS
Polyoma -340	AGGAAGTGACT	HS,H
SFFV ⁶ LTR -110	AGGACCTGAAA	HS
CONSENSUS	VGGAAGTGAV	HS

¹fettgedruckte Buchstaben der Literaturhinweise²major late promoter³avian sarcoma virus - long terminal repeat⁴BK Virus, ein menschliches Papovavirus⁵mouse mammary tumor virus⁶murine Friend spleen focus forming virus

7.2.2.

SV40 Enhancer Stammsequenzen

(Weiher et al., 1983; Gillies et al., 1983; Hen et al. 1983; Herbolmel et al., 1984; Mosthaf et al., 1985; Boshart et al., 1985)

Ad12 157	TGTGACGT	H3
Ad2 158	TGTGGTAA	H3, H4
Ad7 170	TGCGGTAA	H3
BKV	GGACTGGTTTG	W
BPV ⁷ repeat 1	TGTGGACC	H3
BPV repeat 2	TGTGTACC	H3
CMV ⁸ 4x 18 repeat	TGGAAAGTCCC	B
IgH1	TGTGGTTTG	G, H3, H4
IgH2	GTGGTTTTG	G, H4
IgH3	TGTGGCAAG	H3
IgK1	TGGTAAAG	G
IgK2	TCGGAAG	G
IgL	TGGAAG	G
MSV ⁹	TGTGGTAAG	W, G, H3, H4
Polyoma 5120	GTGTGGTTTTG	W, G, H3, H4
LPV ¹⁰	GCTGTGGTTAG	M
SV40 178	GGTGTGGAAG	W, G, H3, H4

⁷bovine papilloma virus

⁸human cytomegalo virus

⁹murine sarcoma virus

¹⁰lymphotropic papovavirus

7.2.3.

SV40 SphI Motive

(Parslow et al., 1984; Falkner & Zachau, 1984; Bergman et al., 1984; Krol et al., 1985; Ciliberto et al., 1985; Mattaj et al., 1985; Zenke et al., 1986)

dc-Box, consensus	ATGCAAATCA	FZ
Ig Pro/Enh; Ulb/U2	ATGCAAAT	P, FZ, B, K, C, M
Ula	ATGTAAAC	K, C
Ul.1	ATGTAGAT	K, C
Ul.3	ATGTAAAT	K
Ig Promotor	ATGCAAAG	FZ
SV40 140	ATGCAAAGCATGCAT	
LPV	TGCA AGTCTGCA	Z
Polyoma 5195	TGCA AG	

7.2.4.

TCCCCA Motive

(Grosschedl et al., 1983; Melton et al., 1984; Zenke et al., 1986)

SV40 168	GTCCCCAGGCTCCCCAGC	Z
HPRT ¹¹ Pro rep	CCAGGCTCC	M
SSV ¹²	GTTCCCAGA	G
MSV/SFFV	GTCCCCAGA	G

¹¹Hypoxanthin Phosphoribosyltransferase

¹²simian sarcoma virus

7.2.5.

Histon Gen Enhancerelemente

(Grosschedl et al., 1983)

H2A -129	AACAGATGG
H2A -148	GCTGCTTCA
H2A -154	GCTGCTCTA
SFFV/MSV/H2A/H3	AACAGATGG
SSV -188	GACAGATGG

7.2.6.

Regulatorische Elemente der hsp Gene

(Pelham, 1982; Wu, 1984; Hultmark et al., 1986; Wu et al., 1986)

HSE consensus	CTNGAANN TTCNNG	P
HSE hsp70	CTCGAATG TTCGCG	W4
HSE hsp83	CTCTAGAAGTTTCTAGAG	W4
hsp22 cap	CAGTTCAAA	H
hsp23 cap	CAGTTGAAT	H
hsp26 cap	CAGATCGAA	H
hsp27 cap	CAGTCTAAA	H
hsp68 cap	CATTTGAAA	H
hsp70 cap	CAATTCAAT	H
hsp70 3' zu TATA	AGAGGCGC	W4
hsp83 3' zu TATA	GCAGACGC	W4
hsp70 SRE ¹³	GAAGGGAAAAAG	W6

¹³serum responsible element

7.2.7.

Weitere Sequenzelemente

(Proudfoot & Brownlee, 1976; Benoist et al., 1980; Fowlkes & Shenk, 1980; Corden et al., 1980; Bogenhagen & Brown, 1981; Nagata et al., 1983; Whitton & Clements, 1984; Borgmeyer et al., 1984; Gronostajski et al., 1985; Carthew et al., 1985; Kadonaga et al., 1986; Jones et al., 1986; G. Antoine, persönliche Mitteilung)

Ad2 MLP bind. Prot.	GGTGTAGGCCACGTGACC	Ca
Ad2/Ad5 ITR NF1 ¹⁴	TTGGATTGAAGCCAA	N
AdFL ¹⁵ ITR NF1	TTGGCTTTGTTCCAA	A
AdFL re. Ende NF1	TTGGCACTGTGCCAA	A (Pos. 123)
NF1 consensus	TGG N ₆₋₇ GCCAA	G
TGGCA bind. Protein	QTGGCANNNTGCCAP	Bo
CAAT Box	GGQTCATCT	Be
HSV IE Element	TAATGARAT	WC
Poly-A Signal	AATAAA	PB
PolIII 5' Transkr. Signal	GTGGQNNPGTGG	FS
PolIII 3' Transkr. Signal	GGGTTCGAANCC	FS
PolIII Terminator	GCTTTTGC	BB
Spl core	GGGCGG	K
Spl consensus	WGGGCGGPPQ	K
Spl Box I, HTLV-III ¹⁶	GGGGAGTGGC	J
Spl Box III, HTLV-III	GAGGCGTGGC	J
TATA Box	gTATARAR	Co

¹⁴nuclear factor I (binding site)

¹⁵mouse adenovirus type FL

¹⁶human T-lymphotropic virus type III (= ARV-2, LAV, "AIDS-virus", HIV)

7.2.8.

Spleiß Signale

(Sakano et al., 1980; Mount, 1982; Siu et al., 1984; Kavalier et al., 1984; Ruskin et al., 1984; Konarska et al., 1985)

5' Donor	VAGGTPAGT	M
3' Acceptor	Q ₁₁ NQAGG	M
Verzweigungsstelle	QNQT <u>PAQ</u>	R (branch site)
Ad2 MLP, Verzw'sstelle	TACTTATCC	Ko
Switch 4+5	GGTAN ₁₋₃ CAPAG	Sa
Switch 5	ATGAG	Sa
Switch 5'	ATGGG	Sa
Switch 5''	CTAAA	Sa
Switch 5 ³ '	CTAAG	Sa
Switch 5 ⁴ '	CTAGG	Sa
Switch 5 ⁵ '	CTGAG	Sa
Switch 5 ⁶ '	CTGGA	Sa
Switch 5 ⁷ '	CTGGG	Sa
Switch 5 ⁸ '	CTGAT	Sa
Switch 5 ⁹ '	TTGAG	Sa
Switch 6	TCCTGG	Sa
Switch 6+3	TCCTGGN ₇₋₉ AGA	Sa
VJ7 Ig cons.	CACTGTG	Sa
VJ 7K2	CAGTGTG	Sa
VJ 7K3	CACTGTA	Sa
VJ 7B cons.	TGCTGTG	Ka
VJ 7B1-2	TTATGTG	Si
VJ 7B2	CGCTGTG	Si
VJ 7H1	GA CTGTG	Sa
VJ7 H2	TAGTGTG	Sa
VJ7 H3	CAATGTG	Sa
VJ7 H4	TATTGTG	Sa
VJ7 L1	CATTGTG	Sa

VJ9 Ig cons.	GATTTTTGT	Sa
VJ9 B2	GCAGAAACC	Ka
VJ9 H1	AGTTTTAGT	Sa
VJ9 H1'	ACAAATACC	Sa
VJ9 H3	ATTTATTGT	Sa
VJ9 K2	AGTTTTTGT	Sa
VJ9 K2'	ACATAAACC	Sa
VJ9 K3	GGGTTTTGT	Sa
VJ9 L1	GGTTTTTGC	Sa
VJ9 L1'	TCAAGAACA	Sa
VJ9 L2	GGTTTTGGG	Si
VJ9 L2'	ACAAGAACA	Sa
VJ9 L3	GGTTTAGGG	Si

7.2.9.

S1 Nuklease hypersensitive Stellen

(Evans et al., 1984)

chi β -globin	CGGGGAAGAG
hu β -globin	CAGGGCAGAG
hu β -globin -200	CCAAGAGA
rat insulin	CAGGGGTCA
SV40 151	GCTGGGGAG
SV40 5212	CACTACTTCTG

7.3.

Alternierende P/Q Sequenzen

p9, 529bp				TC7p9, 1132bp			pCaOri7.01, 436bp		
2	44.5%	43.9%	1.01	49.9%	40.5%	1.23	48.4%	43.0%	1.13
5	26	24	1.1	71	33	2.1	25	11	2.3
6	12	9	1.3	35	17	2.1	12	6	2.0
7	5.8	3	1.9	18	8	2.2	6.2	4	1.6
8	2.6	-	-	8.8	4	2.2	3.1	2	1.6
9	1.3	-	-	4.4	2	2.2	1.5	1	1.5
10				2.2	1	2.2			
11				1.1	-	-			

SV40, 5243 bp				Ad12, 1-3957			pBR322, 4363 bp		
2	50.0%	44.4%	1.13	49.9%	48.5%	1.03	50.0%	50.4%	0.99
5	327	185	1.77	247	203	1.22	273	223	1.22
6	164	72	2.3	124	80	1.6	136	101	1.34
7	82	24	3.4	62	34	1.8	68	51	1.3
8	41	5	8.2	31	17	1.8	34	25	1.4
9	20	1	20	15	4	3.8	17	10	1.7
10	10	-	-	7.7	2	3.9	8.5	1	8.5
11				3.9	-	-	4.3	-	-

Tabelle A1: Sequenzen mit alternierenden P/Q

Das Auftreten von Di- (in %) und Oligonukleotiden (Anzahl) mit alternierenden P/Q in den Sequenzen p9, TC7p9, pCaOri7.01, SV40, Ad12 1-3957 (van Ormond & Galibert, 1984) und pBR322 (mittlere Spalten) im Vergleich zum statistisch erwarteten Wert (linke Spalten). Die dritten Spalten geben den Quotienten aus dem statistisch erwarteten und dem tatsächlichen Wert an. Bei der Berechnung des ersten wurde die jeweilige Basenzusammensetzung berücksichtigt.

Dinukleotide mit alternierenden P/Q sind kaum, Oligonukleotide mit alternierenden P/Q sind zum Teil erheblich gegenüber dem statistisch erwarteten Wert reduziert.

Name	Länge bp	P/Q Folgen reduziert
HUMAN BETA GLOBIN REGION ON CHROMOSOME 11: EPSILON GENE	4805	ca. 60%
HUMAN CLASS I TRANSPLANTATION ANTIGENE (HLA) GENE	4123	45-60%
HUMAN (GH) GERMLINE C-MYC PROTO-ONCOGENE, 5' FLANK THROUGH EXON 2	3605	50-60%
HUMAN DIHYDROFOLATE REDUCTASE GENE, EXON 6 AND 3' FLANK	3464	10-20%
CHICKEN CYTOPLASMIC THYMIDINE KINASE GENE, COMPLETE CODING	3008	25-40%
X.LAEVIS EXTERNAL TRANSCRIBED SPACER (ETS) & 18S, 5.8S, 28S, (5'	3924	30-50%
BOVINE MITOCHONDRION, COMPLETE GENOME	16338	ca. 35%
HUMAN MITOCHONDRION, COMPLETE GENOME	16569	40-60%
A.NIDULANS MT OXIA GENE FOR CYTOCHROME OXIDASE SUBUNIT 1	5218	ca. 35%
YEAST (S.CEREVISIAE) ALCOHOLDEHYDROGENASE II GENE, ADR2 (GLUCOSE	3627	30-50%
YEAST (S.CEREVISIAE) CPA II GENE CODING FOR THE ARGININE SPECIFIC	4014	30-40%
YEAST (S.CEREVISIAE) 25S RNA GENE AND FLANKS	3911	40-60%
MAIZE MITOCHONDRIAL 26S RNA GENE AND FLANKS	5218	ca. 35%

weiter nächste Seite

SV40	5243	40-90%
HAMSTER POLYOMA VIRUS (STRAIN A3), COMPLETE GENOME	5295	ca. 30%
ADENOVIRUS TYPE 2, COMPLETE GENOME	35937	ca. 10%
daraus 1-5243		20-50%
ADENOVIRUS TYPE 5: LEFT 32% OF THE GENOME (COORDINATES 0% TO 32.39%	11570	ca. 20%
ADENOVIRUS TYPE 12 LEFT END OF THE GENOME, MAP COORDINATES 0.0 TO	3957	20-45%
EPSTEIN-BARR VIRUS (EBV) B95-85 STRAIN, COMPLETE GENOME	172282	ca. 25%
HUMAN T-CELL LEUKEMIA TYPE III (AIDS; PROVIRAL), ISOLATE BH10	9749	30-60%
AVIAN MYELOCYSTEMATOSIS VIRUS MC29 (PROVIRAL), COMPLETE P96	3779	50-75%
PLASMID PBR322, COMPLETE GENOME	4363	ca. 25%
PLASMID RSC13 (DERIVED FROM R1), COMPLETE GENOME	7891	25-40%
BACTERIOPHAGE F1, COMPLETE GENOME	6407	40-65%
BACTERIOPHAGE LAMBDA, COMPLETE GENOME	48502	30-40%
E. COLI GALR, LYSA & LYS R GENES CODING FOR GALETK OPERON REPRESSOR	4295	10-40%
E. COLI REPLICATION ORIGIN (ORIC) & ASNA GENE CODING FOR; 2	4012	0-15%
S. TYPHIMURIUM ARGCT GENE & HISTIDINE TRANSPORT OPERON	4483	0-5%
<i>Weiter nächste Seite</i>		

SV40, untranscribed region

414 3-27%

Ad2, 1-560

----- +10%

Ad12, 1-502

----- +30%

HUMAN CYTOMEGALOVIRUS IMMEDIATE EARLY (IE) GENE 1, PROMOTOR-

490

----- +20%

HERPES SIMPLEX VIRUS TYPE 1 SHORT REGION, TR-S/U-S BOUNDARY

4827

----- +15%

Tabelle A2: Sequenzen mit alternierenden P/Q in der DNA von Eukaryoten, Viren, Prokaryoten und Phagen

Dazu wurden aus der MicroGenie™ Datenbank ca. 4 kb lange Sequenzen aus eukaryotischen und prokaryotischen Genomen und möglichst komplette Sequenzen der DNA von Mitochondrien, Viren Phagen und Plasmiden zufällig ausgewählt und nach alternierenden P/Q Folgen (5bp und länger) abgesucht. Die Anzahl der tatsächlich vorkommenden Folgen von alternierenden P/Q aller Längen wurde mit den statistisch zu erwartenden Werten, bei deren Berechnung die spezifischen Basenzusammensetzungen berücksichtigt wurden, verglichen. Spalte 3 gibt an, um wieviel die tatsächlich vorkommenden alternierenden P/Q Folgen gegenüber den statistisch erwarteten Werten reduziert sind. Die mit Großbuchstaben geschriebenen Bezeichnungen sind exakt die Namen, unter denen die Sequenzen in der MicroGenie™ Datenbank zu finden sind.

7.4.

Statistische Signifikanz von Sequenzhomologien

(Kanehisa, 1984)

$$P(m) = p^m$$

$$P(m, r) = \frac{(m-2+r)! p^m (1-p)^r}{(m-2)! r!}$$

$$P(m, r, g) = \frac{(m-2+r+g)! p^m (1-p)^r}{(m-2)! r! g!}$$

P = Wahrscheinlichkeit für das Auftreten einer bestimmten Sequenzhomologie

m = matches (passende Nukleotide)

r = mismatches (nicht passende Nukleotide)

g = gap (ungepaartes Nukleotid; Lücke in der Homologie)

p = average probability of a match (Häufigkeit eines passenden Nukleotids, z.B. je 25% A, C, G und T)

Eine Homologie beginnt und endet gewöhnlich mit einem passenden Nukleotid. Diese Positionen stehen für die Permutation von r und g nicht zur Verfügung und werden deshalb von der Länge der homologen Sequenz abgezogen.

Beispiele:

$$10/10 \text{ Homologie: } P = 0.25^{10} = 9.54 \times 10^{-7}$$

9/10 Homologie, ein "falsches" Nukleotid:

$$P(m=9, r=1) = \frac{(9-2+1)! \cdot 0.25^9 \cdot 0.75}{(9-2)!} = 2.29 \times 10^{-5}$$

9/10 Homologie, ein ungepaartes Nukleotid:

$$P(m=9, g=1) = \frac{(9-2+1)! \cdot 0.25^9}{(9-2)!} = 3.04 \times 10^{-5}$$

8/10 Homologie, ein "falsches", ein ungepaartes Nukleotid:

$$P(m=8, p=1, g=1) = \frac{(8-2+1+1)! \cdot 0.25^8 \cdot 0.75}{(8-2)!} = 6.41 \times 10^{-4}$$

2 Homologien von Kapitel 4.3.7.1.:

$$P(23, 1, 1) = \frac{(23-2+1+1)! \cdot 0.25^{23} \cdot 0.75}{(23-2)!} = 5.4 \times 10^{-12}$$

$$P(22, 3, 2) = \frac{(22-2+3+2)! \cdot 0.25^{22} \cdot 0.75^3}{(22-2)! \cdot 3! \cdot 2!} = 4.8 \times 10^{-9}$$

$m+r$	$P(m)$	$P(m, r=1)$	$P(m) +$ $P(m, r=1)$	$P(m, r=2)$
7	6.10×10^{-5}	9.16×10^{-4}	9.77×10^{-4}	5.49×10^{-3}
8	1.53×10^{-5}	2.75×10^{-4}	2.90×10^{-4}	2.06×10^{-3}
9	3.81×10^{-6}	8.01×10^{-5}	8.39×10^{-5}	7.21×10^{-4}
10	9.54×10^{-7}	2.29×10^{-5}	2.83×10^{-5}	2.40×10^{-4}
11	2.38×10^{-7}	6.44×10^{-6}	6.68×10^{-6}	7.72×10^{-5}
12	5.96×10^{-8}	1.79×10^{-6}	1.84×10^{-6}	2.41×10^{-5}
13	1.49×10^{-8}	4.92×10^{-7}	5.07×10^{-7}	7.38×10^{-6}
14	3.73×10^{-9}	1.34×10^{-7}	1.38×10^{-7}	2.21×10^{-6}
15	9.31×10^{-10}	3.63×10^{-8}	3.73×10^{-8}	6.54×10^{-7}
16	2.33×10^{-10}	9.78×10^{-8}	1.00×10^{-8}	1.91×10^{-7}
17	5.82×10^{-11}	2.62×10^{-9}	2.68×10^{-9}	5.50×10^{-8}
18	1.46×10^{-11}	6.98×10^{-10}	7.13×10^{-10}	1.57×10^{-8}
19	3.64×10^{-12}	1.86×10^{-10}	1.89×10^{-10}	4.45×10^{-9}
20	9.09×10^{-13}	4.91×10^{-11}	5.00×10^{-11}	1.25×10^{-9}
21	2.27×10^{-13}	1.30×10^{-11}	1.32×10^{-11}	3.50×10^{-10}
22	5.68×10^{-14}	3.41×10^{-12}	3.47×10^{-12}	9.72×10^{-11}
23	1.42×10^{-14}	8.95×10^{-13}	9.09×10^{-13}	2.69×10^{-11}
24	3.55×10^{-15}	2.34×10^{-13}	2.38×10^{-13}	7.39×10^{-12}
25				2.02×10^{-12}
26				5.52×10^{-13}
27				1.50×10^{-13}

$$P(m, g=1) = 4/3 P(m, r=1)$$

8.

LITERATURHINWEISE

- Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D. & Brinster, R.L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* 318, 533-538.
- Alitalo, K., Bishop, J.M., Smith, D.H., Chen, E.Y., Colby, W.W. & Levinson, A.D. (1983). Nucleotide sequence of the v-myc oncogene of avian retrovirus MC29. *Proc. Natl. Acad. Sci. USA* 80, 100-104.
- ar-Rushdi, A., Nishikura, K., Erikson, J., Watt, R., Rovera, G. & Croce, C.M. (1983). Differential expression of the translocated and the untranslocated c-myc oncogene in Burkitt lymphoma. *Science* 222, 390-393.
- Ares, M. Jr., Mangin, M. & Weiner, A.M. (1985). Orientation-dependent transcriptional activator upstream of a human U2 snRNA gene. *Mol. Cell. Biol.* 5, 1560-1570.
- Arnott, S. & Selsing, E. (1974). Structures for the polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT). *J. Mol. Biol.* 88, 509-521.
- Azorin, F. & Rich, A. (1985). Isolation of Z-DNA binding proteins from SV40 minichromosomes: evidence for binding to the viral control region. *Cell* 41, 365-374.
- Bach, R., Allet, B. & Crippa, M. (1981). Sequence organization of the spacer in the ribosomal genes of *Xenopus clivii* and *Xenopus borealis*. *Nucl. Acids Res.* 9, 5311-5330.
- Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S. & Barrell, B.G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 310, 207-211.

- Banerji, J., Olson, L. & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33, 729-740.
- Batley, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. & Leder, P. (1983). The human c-myc oncogene: structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* 34, 779-787.
- Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischer, E. & Goodman, H.M. (1980). Sequence of the human insulin gene. *Nature* 284, 26-32.
- Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). The ovalbumin gene - sequence of putative control regions. *Nucl. Acids Res.* 8, 127-142.
- Bentley, D.L. & Rabbits, T.H. (1983). Evolution of immunoglobulin V genes: evidence indicating that recently duplicated Human V κ sequences have diverged by gene conversion. *Cell* 32, 181-189.
- Bergman, Y., Rice, D., Grosschedl, R. & Baltimore, D. (1984). Two regulatory elements for immunoglobulin kappa light chain gene expression. *Proc. Natl. Acad. Sci. USA* 81, 7041-7045.
- Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J. (1983). Sequence of the murine and human cellular myc oncogenes and two modes of myc transcription resulting from chromosome translocation in B lymphoid tumours. *EMBO J.* 2, 2375-2383.
- Biggin, M.D., Gibson, T.J. & Hong, G.F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80, 3963-3965.
- Bishop, J.M. (1983). Cellular oncogenes and retroviruses. *Ann. Rev. Biochem.* 52, 301-354.
- Bishop, J.M. (1985). Viral Oncogenes. *Cell* 42, 23-38.

- Blanchard, J.-M., Piechaczyk, M., Dani, C., Chambard, J.-C., Franchi, A., Pouyssegur, J. & Jeanteur, P. (1985). c-myc gene is transcribed at high rate in Go-arrested fibroblasts and is post-transcriptionally regulated in response to growth factors. *Nature* 317, 443-445.
- Bogenhagen, D.F. & Brown, D.D. (1981). Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. *Cell* 24, 261-270.
- Borgmeyer, U., Nowock, J. & Sippel, A.E. (1984). The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. *Nucl. Acids Res.* 12, 4295-4311.
- Boshart, M., Weber, F., Jahn, G., Dorsch-Häsler, K., Fleckenstein, B. & Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41, 521-530.
- Boyer, H.W. & Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 41, 459-472.
- Breathnach, R. & Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50, 349-383.
- Buchman, A.R., Burnett, L. & Berg, P. (1981). The SV40 nucleotide sequence. In J. Tooze (ed.): "DNA Tumor Viruses", 2nd edition, pp. 799-841, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Butnick, N.C., Miyamoto, C., Chizzonite, R., Cullen, B.R., Ju, G. & Skalka, A.M. (1985). Regulation of the human c-myc gene: 5' noncoding sequences do not affect translation. *Mol. Cell. Biol.* 5, 3009-3016.
- Calladine, C.R. (1982). Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* 161, 343-352.
- Campisi, J., Gray, H.E. Pardee, A.B., Dean, M. & Sonenshein, G.E. (1984). Cell-cycle control of c-myc but not c-ras

- expression is lost following chemical transformation. *Cell* 36, 241-247.
- Carthew, R.W., Chodosh, L.A. & Sharp, P.A. (1985). A RNA polymerase transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell* 43, 439-448.
- Chen, M.-J., Shimada, T., Moulton, A.D., Cline, A., Humphries, R.K., Maizel, J. & Nienhuis, A.W. (1984). The functional human dihydrofolate reductase gene. *J. Biol. Chem.* 259, 3933-3943.
- Chen, E.J. & Seeburg, P.H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* 4, 165-170.
- Chien, Y., Gascoigne, N.R.J., Kavalier, J., Lee, N.E. & Davis, M.M. (1984). Somatic recombination in a murine T-cell receptor gene. *Nature* 309, 322-326.
- Ciliberto, G., Buckland, R., Cortese, R. & Philipson, L. (1985). Transcriptional signals in embryonic *Xenopus laevis* U1 RNA genes. *EMBO J.* 4, 1537-1543.
- Clarke, L. & Carbon, J. (1976). A colony bank containing synthetic Col EI hybrid plasmids representative of the entire *E. coli* genome. *Cell* 9, 91-99.
- Colby, W.W., Chen, E.Y., Smith, D.H. & Levinson, A.D. (1983). Identification and nucleotide sequence of a human locus homologous to the v-myc oncogene of avian myelocytomatosis virus MC29. *Nature* 301, 722-725.
- Cole, M.D. (1985). Regulation and activation of c-myc. *Nature* 318, 510-511.
- Collins, S. & Groudine, M. (1982). Amplification of endogenous myc-related DNA sequences in a human myeloid leukemia cell line. *Nature* 298, 679-681.
- Conrad, S.E. & Botchan, M.R. (1982). Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. *Mol. Cell. Biol.* 2, 949-965.

- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. & Chambon, P. (1980). Expression of cloned genes in new environment. *Science* **209**, 1406-1414.
- Cordingley, M.G., Campbell, M.E.M. & Preston, C.M. (1983). Functional analysis of a herpes simplex virus type 1 promoter: identification of far-upstream regulatory sequences. *Nucl. Acids Res.* **11**, 2347-2365.
- Crews, S., Barth, R., Hood, L., Prehn, J. & Calame, K. (1982). Mouse c-myc oncogene is located on chromosome 15 and translocated to chromosome 12 in plasmacytomas. *Science* **218**, 1319-1321.
- Dalla-Favera, R., Gelmann, E.P., Martinotti, S., Franchini, G., Papas, T.S., Gallo, R.C. & Wong-Stahl, F. (1982). Cloning and characterization of different human sequences related to the onc gene (v-myc) of avian myelocytomatosis virus (MC29). *Proc. Natl. Acad. Sci. USA* **79**, 6497-6501.
- Dani, C., Blanchard, J.M., Piechaczyk, M., El Sabouty, S., Marty, L. & Jeanteur, P. (1984). Extreme instability of myc mRNA in normal and transformed human cells. *Proc. Natl. Acad. Sci. USA* **81**, 7046-7050.
- Dani, C., Mechti, N., Piechaczyk, M., Lebleu, B., Jeanteur, P. & Blanchard, J.M. (1985). Increased rate of degradation of c-myc mRNA in interferon-treated Daudi cells. *Proc. Natl. Acad. Sci. USA* **82**, 4896-4899.
- Darveau, A., Pelletier, J. & Sonenberg, N. (1985). Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. USA* **82**, 2315-2319.
- Davison, A.J. & Wilkie, N.M. (1981). Nucleotide sequence of the joint between the L and S segments of herpes simplex virus types 1 and 2. *J. gen. Virol.* **55**, 315-331.
- DeLucia, A.L., Lewton, B.A., Tjian, R. & Tegtmeyer, P. (1983). Topography of simian virus 40 A Protein-DNA complexes:

- arrangement of pentanucleotide interaction sites at the origin of replication. *J. Virol.* 46, 143-150.
- Dhar, R., McClements, W.L., Enquist, L.W. & Vande Woude, G.F. (1980). Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc. Natl. Acad. Sci. USA* 77, 3937-3941.
- Donner, P., Greiser-Wilke, I. & Moelling, K. (1982). Nuclear localisation and DNA binding of the transforming gene product of avian myelocytomatosis virus. *Nature* 296, 262-266.
- Dony, C., Kessel, M. & Gruss, P. (1985). Post-transcriptional control of *myc* and *p53* expression during differentiation of the embryonal carcinoma cell line F9. *Nature* 317, 636-639.
- Dulbecco, R. & Freeman, G. (1959). Plaque production by the polyoma virus. *Virology* 8, 396-397.
- Dynan, W.S. & Tjian, R. (1983a). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 32, 669-680.
- Dynan, W.S. & Tjian, R. (1983b). The promoter-specific transcription Factor Spl binds to upstream sequences in the SV40 early promoter. *Cell* 35, 79-87.
- Dynan, W.S., Saffer, J.D., Lee, W.S. & Tjian, R. (1985). Transcription factor Spl recognizes promoter sequences from the monkey genome that are similar to the simian virus 40 promoter. *Proc. Natl. Acad. Sci. USA* 82, 4915-4919.
- Dynan, W.S., Sazer, S., Tjian, R. & Schimke, R.T. (1986). Transcription factor Spl recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature* 319, 246-248.
- Elgin, S.C.R. (1984). Anatomy of hypersensitive sites. *Nature* 309, 213-214.

- Ellison, J.W., Berson, B.J. & Hood, L.E. (1982). The nucleotide sequence of a human immunoglobulin C γ gene. *Nucl. Acids Res.* 10, 4071-4079.
- Erikson, J., Finger, L., Sun, L., ar-Rushsdi, A., Nishikura, K., Minowada, J., Finan, J., Emanuel, B.S., Nowell, P.C. & Croce, C.M. (1986). Deregulation of c-myc by translocation of the α -locus of the T-cell receptor in T-cell leukemias. *Science* 232, 884-886.
- Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. (1984). SI-hypersensitive sites in eukaryotic promoter regions. *Nucl. Acids Res.* 12, 8043-8057.
- Falkner, F.G. & Zachau, H.G. (1984). Correct transcription of an immunoglobulin kappa gene requires an upstream fragment containing conserved sequence elements. *Nature* 310, 71-74.
- Fowlkes, D.M. & Shenk, T. (1980). Transcriptional control regions of the adenovirus VAI RNA gene. *Cell* 22, 405-413.
- Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J. Mol. Biol.* 170, 827-842.
- Fütterer, J. & Winnacker, E.-L. (1984). Adenovirus DNA replication. *Current Topics in Micro. and Immunol.* 111, 41-64.
- Gazin, C., Dupont de Dinechin, S., Hampe, A., Masson, J.-M., Martin, P., Stehelin, D. & Galibert, F. (1984). Nucleotide sequence of the human c-myc locus: provocative open reading frame within the first exon. *EMBO J.* 3, 383-387.
- Gidoni, D., Dynan, W.S. & Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* 312, 409-413.
- Gidoni, D., Kadonaga, J.T., Barera-Saldana, H., Takahashi, K., Chambon, P. & Tjian, R. (1985). Bidirectional SV40

- transcription mediated by tandem Spl binding interactions. *Science* 230, 511-517.
- Gillies, S.D., Morrison, S.L., Oi, V.T. & Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33, 717-728.
- Gorman, C.M., Moffat, L.F. & Howard, B.H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2, 1044-1051.
- Griffin, B.E., Soeda, E., Barrell, B.G. & Staden, R. (1981). Sequence and analysis of polyoma virus DNA. In J. Tooze (ed.): "DNA Tumor Viruses", 2nd edition, pp. 843-910, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Gronostajski, R.M., Adhya, S., Nagata, K., Guggenheimer, R.A. & Hurwitz, J. (1985). Site-specific DNA binding of nuclear factor I: analyses of cellular binding sites *Mol. Cell. Biol.* 5, 964-971.
- Grosschedl, R., Mächler, M., Rohrer, U. & Birnstiel, M.L. (1983). A functional component of the sea urchin H2A gene modulator contains an extended sequence homology to a viral enhancer. *Nucl. Acids Res.* 11, 8123-8136.
- Grosschedl, R. & Baltimore, D. (1985). Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* 41, 885-897.
- Hall, L.M.C. & Maden, B.E.H. (1980). Nucleotide sequence through the 18S-28S intergene region of a vertebrate ribosomal transcription unit. *Nucl. Acids Res.* 8, 5993-6005.
- Hammarskjöld, M.-L. & Winberg, G. (1980). Encapsidation of adenovirus 16 DNA is directed by a small DNA sequence at the left end of the genome. *Cell* 20, 787-795.
- Hann, S.R., Abrams, H.D., Rohrschneider, L.R. & Eisenmann, R.N. (1983). Proteins encoded by v-myc and c-myc oncogenes: identification and localisation in acute leukemia virus

- transformants and bursal lymphoma cell lines. *Cell* 34, 789-798.
- Harada, F. & Kato, N. (1980). Nucleotide sequence of 4.5S RNAs associated with poly (A)-containing RNAs of mouse and hamster cells. *Nucl. Acids Res.* 8, 1273-1285.
- Hart, R.P. & Folk, W.R. (1982). Structure and organization of a mammalian 5S gene cluster. *J. Biol. Chem.* 257, 11706-11711.
- Haynes, S.R., Toomey, T.P., Leinwand, L. & Jelinek, W.R. (1981). The chinese hamster Alu-equivalent sequence: a conserved highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. *Mol. Cell. Biol.* 1, 573-583.
- Hearing, P. & Shenk, T. (1983). The adenovirus type 5 ElA transcriptional control region contains a duplicated enhancer element. *Cell* 33, 695-703.
- Heilig, R., Muraskowsky, R. & Mandel, J.-L. (1982). The ovalbumin gene family: the 5' end region of the X and Y genes. *J. Mol. Biol.* 156, 1-19.
- Hen, R., Borrelli, E., Sassone-Corsi, P. & Chambon, P. (1983). An enhancer element is located 340 base pairs upstream from the adenovirus-2 ElA capsite. *Nucl. Acids Res.* 11, 8747-8760.
- Hendrix, R.W., Roberts, J.W., Stahl, F.W. & Weisberg, R.A. (1983). Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Herbomel, P., Bourachot, B. & Yaniv, M. (1984). Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell* 39, 653-662.
- Hieter, P.A., Max, E.E., Seidman, J.G., Maizel, J.V. Jr. & Leder, P. (1980). Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. *Cell* 22, 197-207.

- Hinnebusch, A.G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci. USA* 81, 6442-6446.
- Hohn, B. (1979). *In vitro* packaging of Lambda and cosmid DNA. *Methods Enzymol.* 68, 299-309.
- Hoogsteen, K. (1959). The structure of crystals containing a hydrogen bonded complex of 1-methylthymine and 9-methyladenine. *Acta Cryst.* 12, 822-823.
- Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst.* 16, 907-916.
- Houck, C.M., Rinehart, F.P. & Schmid, C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.* 132, 289-306.
- Hultmark, D., Klemenz, R. & Gehring, W.J. (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell* 44, 429-438.
- Hunt, C. & Morimoto, R.I. (1985). Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. *Proc. Natl. Acad. Sci. USA* 82, 6455-6459.
- Imperiale, M.J., Feldman, L.T. & Nevins, J.R. (1983). Activation of gene expression by adenovirus and herpesvirus regulatory genes acting in trans and by a cis-acting adenovirus enhancer element. *Cell* 35, 127-136.
- IUPAC (1971). Tentative rules for the nomenclature of organic chemistry. *Eur. J. Biochem.* 18, 151-170.
- Jensen, F.C., Girardi, A.J., Gilden, R.V. & Koprowski, H. (1964). Infection of human and simian tissue cultures with rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* 52, 53-59.
- Jones, K.A. & Tjian, R. (1984). Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. *Cell* 36, 155-162.

- Jones, K.A. & Tjian, R. (1985). Spl binds to promoter sequences and activates herpes simplex virus 'immediate early' gene transcription *in vitro*. *Nature* 317, 179-182.
- Jones, K.A., Yamamoto, K.R. & Tjian, R. (1985). Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. *Cell* 42, 559-572.
- Jones, K.A., Kadonaga, J.T., Luciw, P.A. & Tjian, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Spl. *Science* 232, 755-759.
- Kadonaga, J.T., Jones, K.A. & Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Spl. *Trends Biochem. Sci.* 11, 20-23.
- Kanehisa, M. (1984). Use of statistical criteria for screening potential homologies in nucleic acid sequences. *Nucl. Acids Res.* 12, 203-213.
- Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980). Novel bacteriophage Lambda cloning vector. *Proc. Natl. Acad. Sci. USA* 77, 5172-5176.
- Karn, J., Matthes, H.W.D., Gait, M.J. & Brenner, S. (1984). A new selective cloning vector, Lambda 2001, with sites for XbaI, BamHI, HindIII, EcoRI, SstI and XhoI. *Gene* 32, 217-224.
- Kavaler, J., Davis, M.M. & Chien, Y. (1984). Localization of a T-cell receptor diversity-region element. *Nature* 310, 421-423.
- Keath, E.J., Kelekar, A. & Cole, M.D. (1984). Transcriptional activation of the translocated c-myc oncogene in mouse plasmacytomas: similar RNA levels in tumor and proliferating normal cells. *Cell* 37, 521-528.
- Kelly, K., Cochran, B.H., Stiles, C.D. & Leder, P. (1983). Cell-specific regulation of the c-myc gene by lymphocyte mitogenes and platelet-derived growth factor. *Cell* 35, 603-610.

- Kingston, R.E., Baldwin, A.S. Jr., & Sharp, P.A. (1984). Regulation of heat shock protein 70 gene expression by c-myc. *Nature* 312, 280-282.
- Kingston, R.E., Baldwin, A.S. & Sharp, P.A. (1985). Transcription control by oncogenes. *Cell* 41, 3-5.
- Klein, G. & Klein, E. (1985). Evolution of tumours and the impact of molecular oncology. *Nature* 315, 190-195.
- Kohl, N.E., Kanda, N., Schreck, R.R., Bruns, G., Latt, S.A., Gilbert, F. & Alt, F.W. (1983). Transposition and amplification of oncogene related sequences in human neuroblastomas. *Cell* 35, 359-367.
- Kohl, N.E., Legouy, E., DePinho, R.A., Nisen, P.D., Smith, R.D., Gee, C.E. & Alt, F.W. (1986). Human N-myc is closely related in organization and nucleotide sequence to c-myc. *Nature* 319, 73-77.
- Konarska, M.M., Grabowski, P.J., Padgett, R.A. & Sharp, P.A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* 313, 552-557.
- Kost, T.A., Theodorakis, N. & Hughes, S.H. (1983). The nucleotide sequence of the chick cytoplasmic β -actin gene. *Nucl. Acids Res.* 11, 8287-8301.
- Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl. Acids Res.* 12, 857-872.
- Krol, A., Lund, E. & Dahlberg, J.E. (1985). The two embryonic U1 RNA genes of *Xenopus laevis* have both common and gene-specific transcription signals. *EMBO J.* 4, 1529-1535.
- Kruczek, I. & Doerfler, W. (1983). Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: effect of promoter methylation on gene expression. *Proc. Natl. Acad. Sci. USA* 80, 7586-7590.

- Kwoh, T.J. & Engler, J.A. (1984). The nucleotide sequence of the chicken thymidine kinase gene and the relationship of its predicted polypeptide to that of the vaccinia virus thymidine kinase. *Nucl. Acids Res.* 121, 3959-3971.
- Land, H., Parada, L.F. & Weinberg, R.A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* 304, 596-602.
- Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T. & Taub, R. (1983). Translocation among antibody genes in human cancer. *Science* 222, 765-771.
- Levine, A.J. (1984). The adenovirus early proteins. *Current Topics in Micro. and Immunol.* 110, 143-167.
- Linial, M. & Groudine, M. (1985). Transcription of three c-myc exons is enhanced in chicken bursal lymphoma cell lines. *Proc. Natl. Acad. Sci. USA* 82, 53-57.
- Logan, J. & Shenk, T. (1984). Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proc. Natl. Acad. Sci. USA* 81, 3655-3659.
- Maden, B.E.H., Moss, M. & Salim, M. (1982). Nucleotide sequence of an external transcribed spacer in *Xenopus laevis* rDNA: sequences flanking the 5' and 3' ends of 18S rRNA are non-complementary. *Nucl. Acids Res.* 10, 2387-2398.
- Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Margolskee, R.F. & Nathans, D. (1984). Simian virus 40 mutant T antigens with relaxed specificity for the nucleotide sequence at the viral DNA origin of replication. *J. Virol.* 49, 386-393.
- Mason, J.O., Williams, G.T. & Neuberger, M.S. (1985). Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. *Cell* 41, 479-487.

- Mattaj, I.W., Lienhard, S., Jiricny, J. & De Robertis, E.M. (1985). An enhancer-like sequence within the *Xenopus* U2 gene promoter facilitates the formation of stable transcription complexes. *Nature* 316, 163-167.
- McCutchan, T.F. & Singer, M.F. (1981). DNA sequences similar to those around the simian virus 40 origin of replication are present in the monkey genome. *Proc. Natl. Acad. Sci. USA* 78, 95-99.
- McGarry, T.J. & Lindquist, S. (1985). The preferential translation of *Drosophila* hsp70 mRNA requires sequences in the untranslated leader. *Cell* 42, 903-911.
- Melton, D.W., Konecki, D.S., Brennand, J. & Caskey, C.T. (1984). Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* 81, 2147-2151.
- Messing, J. & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19, 269-276.
- Montell, C., Fisher, E.F., Caruthers, M.H. & Berk, A.J. (1982). Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site. *Nature* 295, 380-384.
- Montminy, M.R., Goodman, R.H., Horovitch, S.J. & Habener, J.F. (1984). Primary structure of the gene encoding rat preprosomatostatin. *Proc. Natl. Acad. Sci. USA* 81, 3337-3340.
- Mosthaf, L., Pawlita, M. & Gruss, P. (1985). A viral enhancer element specifically active in human haematopoietic cells. *Nature* 315, 597-600.
- Mount, S.M. (1982). A catalogue of splice junction sequences. *Nucl. Acids Res.* 10, 459-472.
- Müller, R., Tremblay, J.M., Adamson, E.D. & Verma, I.M. (1983). Tissue and cell type-specific expression of two human c-onc genes. *Nature* 304, 454-456.

- Murphy, J.T., Burgess, R.R., Dahlberg, J.E. & Lund, E. (1982). Transcription of a gene for human U1 small nuclear RNA. *Cell* 29, 265-274.
- Nagata, K., Guggenheimer, R.A. & Hurwitz, J. (1983). Specific binding of a cellular DNA replication protein to the origin of replication of adenovirus DNA. *Proc. Natl. Acad. Sci. USA* 80, 6177-6181.
- Nau, M.M., Brooks, B.J., Battey, J., Sausville, E., Gazdar, A.F., Kirsch, I.R., McBride, O.W., Bertness, V., Hollis, G.F. & Minna, J.D. (1985). L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* 318, 69-73.
- NC-IUB (1985). Nomenclature for incompletely specified bases in nucleic acid sequences. *Eur. J. Biochem.* 150, 1-5.
- Nevins, J.R. (1982). Induction of the synthesis of a 70,000 Dalton mammalian heat shock protein by the adenovirus E1A gene product. *Cell* 29, 913-919.
- Nilsen, T.W. & Maroney, P.A. (1984). Translational efficiency of c-myc mRNA in Burkitt lymphoma cells. *Mol. Cell. Biol.* 4, 2235-2238.
- Nordheim, A. & Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. *Nature* 303, 674-678.
- Norrandner, J., Kempe, T. & Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26, 101-106.
- Nussinov, R. (1980). Some rules in the ordering of nucleotides in the DNA. *Nucl. Acids Res.* 8, 4545-4562.
- Parker, C.S. & Topol, J. (1984). A *Drosophila* RNA polymerase II transcription factor binds to the regulatory site of an hsp70 gene. *Cell* 37, 273-283.
- Parslow, T.G., Blair, D.L., Murphy, W.J. & Granner, D.K. (1984). Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. USA* 81, 2650-2654.

- Pelham, H. (1982). A regulatory upstream promoter element in the *Drosophila* hsp70 heat-shock gene. *Cell* 30, 517-528.
- Pelham, H. (1985). Activation of heat-shock genes in eukaryotes. *Trends Genet.* 1, 31-35.
- Pelletier, J. & Sonenberg, N. (1985). Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 40, 515-526.
- Perry, R.P. (1983). Consequences of myc invasion of immunoglobulin loci: facts and speculation. *Cell* 33, 647-649.
- Persson, H., Henninghausen, L., Taub, R., DeGrado, W. & Leder, P. (1984). Antibodies to human c-myc oncogene product: evidence of an evolutionarily conserved protein induced during cell proliferation. *Science* 225, 687-693.
- Persson, H. & Leder, P. (1984). Nuclear localization and DNA binding properties of a protein expressed by human c-myc oncogene. *Science* 225, 718-721.
- Piechaczyk, M., Yang, J.-Q., Blanchard, J.-M., Jeanteur, P. & Marcu, K.B. (1985). Posttranscriptional mechanisms are responsible for accumulation of truncated c-myc mRNAs in murine plasma cell tumors. *Cell* 42, 589-597.
- Pomerantz, B.J., & Hassel, J.A. (1984). Polyomavirus and simian virus 40 large T antigens bind to common DNA sequences. *J. Virol.* 49, 925-937.
- Proudfoot, N.J. & Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 263, 211-214.
- Proudfoot, N.J., Gil, A. & Maniatis, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* 31, 553-563.
- Queen, C., Lord, S.T., McCutchan, T.F. & Singer, M.F. (1981). Three segments from the monkey genome that hybridize to simian virus 40 have common structural elements. *Mol. Cell. Biol.* 1, 1061-1068.

- Queen, C. & Korn, L.C. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucl. Acids Res.* 12, 581-599.
- Ralston, R. & Bishop, J.M. (1983). The protein products of the myc and myb oncogenes and adenovirus Ela are structurally related. *Nature* 306, 803-806.
- Rich, A., Nordheim, A. & Wang, A.H.-J. (1984). The chemistry and biology of left handed Z-DNA. *Ann. Rev. Biochem.* 53, 791-846.
- Robb, J.A. & Huebner, K. (1973). Effect of chromosome number on simian virus 40 replication. *Exptl. Cell Res.* 81, 120-126.
- Ruskin, B., Krainer, A.R., Maniatis, T. & Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* 38, 317-331.
- Ryder, K., Vakalopoulou, E., Mertz, R., Mastrangelo, I., Hough, P., Tegtmeyer, P. & Fanning, E. (1985). Seventeen base pairs of region I encode a novel tripartite binding signal for SV40 T antigen. *Cell* 42, 539-548.
- Saenger, W. (1983). Principles of nucleic acid structure. Springer Verlag, New York.
- Saffer, J.D. & Singer, M.F. (1984). Transcription from SV40-like monkey DNA sequences. *Nucl. Acids Res.* 12, 4769-4788.
- Saito, H., Hayday, A.C., Wiman, K., Hayward, W.S. & Tonegawa, S. (1983). Activation of the c-myc gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. USA* 80, 7476-7480.
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* 286, 676-683.
- Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

- Sassone-Corsi, P., Hen, R., Borrelli, E., Leff, T. & Chambon, P. (1983). Far upstream sequences are required for efficient transcription from the adenovirus-2 E1A transcription unit. *Nucl. Acids Res.* 11, 8735-8745.
- Schneider, R., Gander, I., Müller, U., Mertz, R. & Winnacker, E.L. (1986). A sensitive and rapid gel retention assay for nuclear factor I and other DNA-binding proteins in crude nuclear extracts. *Nucl. Acids Res.* 14, 1303-1317.
- Scholl, D.R., Kahn, S., Malavarca, R., Astrin, S. & Skalka, A.M. (1983). Nucleotide sequence of the long terminal repeat and flanking cellular sequences of avian endogenous retrovirus ev-2: variation in Rous-associated virus-0 expression cannot be explained by differences in primary sequence. *J. Virol.* 45, 868-871.
- Schubach, W., & Groudine, M. (1984). Alteration of c-myc chromatin structure by avian leukosis virus integration. *Nature* 307, 702-708.
- Schwab, M., Alitalo, K., Klempnauer, K.-H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M. & Trent, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 303, 245-248.
- Scott, M.P. & Pardue, M.L. (1981). Translational control in lysates of *Drosophila melanogaster* cells. *Proc. Natl. Acad. Sci. USA* 78, 3353-3357.
- Seidman, J.G., Max, E.E. & Leder, P. (1979). A kappa-immunoglobulin gene is formed by site specific recombination without further somatic mutation. *Nature* 280, 370-375.
- Seif, I., Khoury, G. & Dhar, R. (1979). The genome of human papovavirus BKV. *Cell* 18, 963-977.
- Serfling, E., Jasin, M. & Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. *Trends Genet.* 1, 224-230.

- Sheiness, D. & Bishop, J.M. (1979). DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. *J. Virol.* 31, 514-521.
- Shih, C.-K., Linial, M., Goodenow, M.M. & Hayward, W.S. (1984). Nucleotide sequence 5' of the chicken c-myc coding region: localisation of a noncoding exon that is absent from myc transcripts in most avian leukosis virus-induced lymphomas. *Proc. Natl. Acad. Sci. USA* 81, 4697-4701.
- Shilo, B.-Z. & Weinberg, R.A. (1981). DNA sequences homologous to vertebrate oncogenes are conserved in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 79, 6789-6792.
- Shima, E.A., Le Beau, M.M., McKeithan, T.W., Minowada, J., Showe, L.C., Mak, T.W., Minden, M.D., Rowley, J.D. & Diaz, M.O. (1986). Gene encoding the α chain of the T-cell receptor is moved immediately downstream of c-myc in a chromosomal 8;14 translocation in a cell line from a human T-cell leukemia. *Proc. Natl. Acad. Sci. USA* 83, 3439-3443.
- Shortle, D. & Nathans, D. (1979). Regulatory mutants of simian virus 40: constructed mutants with base substitutions at the origin of DNA replication. *J. Mol. Biol.* 131, 801-817.
- Siebenlist, U., Henninghausen, L., Battey, J. & Leder, P. (1984). Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma. *Cell* 37, 381-391.
- Singh, H., Sen, R., Baltimore, D. & Sharp, P.A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* 319, 154-158.
- Siu, G., Clark, S.P., Yoshikai, Y., Malissen, M., Yanagi, Y., Strauss, E., Mak, T.W. & Hood, L. (1984). The human T cell antigen receptor is encoded by variable,

- diversity, and joining gene segments that rearrange to generate a complete V gene. *Cell* 37, 393-401.
- Stabel, S., Doerfler, W. & Friis, R.R. (1980). Integration sites of adenovirus type 12 DNA in transformed hamster cells and hamster tumor cells. *J. Virol.* 36, 22-40.
- Stabel, S. & Doerfler, W. (1982). Nucleotide sequence at the site of junction between adenovirus type 12 DNA and repetitive hamster cell DNA in transformed cell line CLACl. *Nucl. Acids Res.* 10, 8007-8023.
- Stanton, L.W., Watt, R. & Marcu, K.B. (1983). Translocation, breakage and truncated transcripts of c-myc oncogene in murine plasmacytomas. *Nature* 303, 401-406.
- Stewart, T.A., Pattengale, P.K. & Leder, P. (1984a). Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. *Cell* 38, 627-637.
- Stewart, T.A., Bellvé, A.R. & Leder, P. (1984b). Transcription and promoter usage of the myc gene in normal somatic and spermatogenic cells. *Science* 226, 707-710.
- Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S. & Honjo, T. (1982). Structure of human immunoglobulin gamma genes: implications for evolution of a gene family. *Cell* 29, 671-679.
- Taub, R., Moulding, C., Battey, J., Murphy, W., Vasicek, T., Lenoir, G. & Leder, P. (1984a). Activation and somatic mutation of the translocated c-myc gene in Burkitt lymphoma cells. *Cell* 36, 339-348.
- Taub, R., Kelly, K., Battey, J., Latt, S., Lenoir, G.M., Tantravahi, U., Tu, Z. & Leder, P. (1984b). A novel alteration in the structure of an activated c-myc gene in a variant t(2;8) Burkitt lymphoma. *Cell* 37, 511-520.
- Thireos, G., Penn, M.D. & Greer, H. (1984). 5' untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl. Acad. Sci. USA* 81, 5096-5100.

- Thomas, P.S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
- Tilghman, S.M., Tiemeier, D.C., Polsky, F., Edgell, M.H., Seidmann, J.G., Leder, A., Enquist, L.W., Norman, B. & Leder, P. (1977). Cloning specific segments of the mammalian genome: bacteriophage Lambda containing mouse globin and surrounding sequences. *Proc. Natl. Acad. Sci. USA* 74, 4406-4410.
- Tjian, R. (1978). The binding site on SV40 DNA for a T antigen-related protein. *Cell* 13, 165-179.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.
- Twigg, A.J. & Sherratt, D. (1980). Trans-complementable copy-number mutants of plasmid ColE1. *Nature* 283, 216-218.
- van der Eb, A.J. & Bernards, R. (1984). Transformation and oncogenicity by adenoviruses. *Current Topics in Micro. and Immunol.* 110, 23-51.
- van Ormond, H. & Galibert, F. (1984). Nucleotide sequences of adenovirus DNAs. *Current Topics in Micro. and Immunol.* 110, 73-142.
- Varmus, H.E. (1984). The molecular genetics of cellular oncogenes. *Ann. Rev. Genet.* 18, 553-612.
- Vieira, J. & Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259-268.
- Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. & Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* 282, 680-686.
- Watson, D.K., Reddy, E.P., Duesberg, P.H. & Papas, T.S. (1983a). Nucleotide sequence analysis of the chicken c-myc gene reveals homologous and unique coding regions by comparison with the transforming gene of avian

- myelocytomatosis virus MC29. *Proc. Natl. Acad. Sci. USA* 80, 2146-2150.
- Watson, D.K., Psallidopoulos, M.C., Samuel, K.P., Dalla-Favera, R. & Papas, T.S. (1983b). Nucleotide sequence analysis of human c-myc locus, chicken homologue, and myelocytomatosis virus MC29 transforming gene reveals a highly conserved gene product. *Proc. Natl. Acad. Sci. USA* 80, 3642-3645.
- Watt, R.A., Shatzman A.R. & Rosenberg, M. (1985). Expression and characterization of the human c-myc DNA-binding protein. *Mol. Cell. Biol.* 5, 448-456.
- Weiher, H., König, M. & Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. *Science* 219, 626-631.
- Weinberg, R.A. (1985). The action of oncogenes in the cytoplasm and nucleus. *Science* 230, 770-776.
- Westin, E.H., Wong-Stahl, F., Gelmann, E.P., Dalla Favera, R., Papas, T.S. Lautenberger, J.A., Eva, A., Reddy, E.P., Tronick, S.R., Aaronson, S.A. & Gallo, R.C. (1982). Expression of retroviral onc genes in human hematopoietic cells. *Proc. Natl. Acad. Sci. USA* 79, 2490-2494.
- Westin, G., Lund, E., Murphy, J., Pettersson, U. & Dahlberg, J. (1984). Human U1 and U2 genes use similar transcription signals. *EMBO J.* 3, 3295-3301.
- Whitton, J.L. & Clements, J.B. (1984). Replication origins and a sequence involved in coordinate induction of the immediate-early gene family are conserved in an inergenic region of herpes simplex virus. *Nucl. Acids Res.* 12, 2061-2079.
- Winnacker, E.-L. (1984). Gene und Klone. Eine Einführung in die Gentechnologie. Verlag Chemie, Weinheim.
- Wright, P.J., DeLucia, A.L. & Tegtmeyer, P. (1984). Sequence-specific binding of simian virus 40 A protein to nonorigin and cellular DNA. *Mol. Cell. Biol.* 4, 2631-2638.

- Wu, B.J. & Morimoto, R. (1985). Transcription of the human hsp70 gene is induced by serum stimulation. *Proc. Natl. Acad. Sci. USA* 82, 6070-6074.
- Wu, B.J., Kingston, R.E. & Morimoto, R. (1986). Human hsp70 promoter contains at least two distinct regulatory domains. *Proc. Natl. Acad. Sci. USA* 83, 629-633.
- Wu, C. (1984). Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. *Nature* 309, 229-234.
- Yancopoulos, G.D., Nisen, P.D., Tesfaye, A., Kohl, N.E., Goldfarb, M.P. & Alt, F.W. (1985). N-myc can cooperate with ras to transform normal cells in culture. *Proc. Natl. Acad. Sci. USA* 82, 5455-5459.
- Zakut, R., Shani, M., Givol, D., Neumann, S., Yaffe, D. & Nudel, U. (1982). Nucleotide sequence of the rat skeletal muscle actin gene. *Nature* 298, 857-859.
- Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. & Chambon, P. (1986). Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* 5, 387-397.
- Zimmerman, J.L., Petri, W. & Meselson, M. (1983). Accumulation of a specific subset of *D.melanogaster* heat shock mRNAs in normal development without heat shock. *Cell* 32, 1161-1170.
- Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.N., Witte, O.N., Toran-Allerand, D., Gee, C.E., Minna, J.D. & Alt, F.W. (1986). Differential expression of myc family genes during murine development. *Nature* 319, 780-783.

Lebenslauf

Name	Herbert Kühnel
Geburtsdatum	7.5.1953
Geburtsort	Immenstadt, Kr. Oberallgäu
Eltern	Josef Kühnel und Maria Kühnel, geb. Geis
Geschwister	5
Familienstand	seit dem 26.5.1982 verheiratet mit Maria Kühnel, geb. Hornung
Kinder	Philipp, * 21.12.1983 Margret, * 30.3.1986

Schulbildung	
1959-1963	Volksschule Thalkirchdorf
1963-1965	Realgymnasium Maria Stern, Immenstadt
1965-1972	Allgäu-Gymnasium, Kempten
23.6.1972	Abitur

Wehrdienst	
1.10.72-30.9.74	Soldat auf Zeit im Fernmeldebataillon 230

Studium	
4.11.1974	Beginn des Chemie Diplom Studiums an der Ludwig-Maximilians-Universität München
28.10.1977	Diplom-Vorprüfung
28.2.1980	Diplom-Hauptprüfung
1.4.80-30.4.81	Diplomarbeit "Untersuchungen an mRNA aus Adenovirus Typ 2 infizierten HeLa Zellen" am Institut für Biochemie unter Anleitung von Prof. Dr. E.-L. Winnacker
1.5.1981-1986	Promotionsstudium, Doktorarbeit unter Anleitung von Prof. Dr. E.-L. Winnacker am Institut für Biochemie der Ludwig- Maximilians-Universität München

